DOCUMENT RESUME

ED 157 240 EC 111 217

AUTHOR Reschly, Daniel J.
TITLE Nonbiased Assessment and School Psychology.
INSTITUTION Iowa State Dept. of Public Instruction, Des Moines. Div. of Pupil Personnel Services.
PUB DATE Aug 78
NOTE 102p.; Appendix may be marginally legible

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Culture Free Tests; Disadvantaged Youth; Elementary Secondary Education; Evaluation Methods; Federal Legislation; *Handicapped Children; *Minority Groups; School Psychologists; *Student Evaluation; *Student Placement; *Test Bias
IDENTIFIERS Education for All Handicapped Chldren Act

ABSTRACT

The paper discusses the various problems related to the current concerns with nonbiased assessment, with particular emphasis on the placement of minority group students in classes for the mildly retarded. Sections address the following topics: special education litigation; underlying issues in court cases (including the meaning of IQ test results, labeling effects, and effectiveness of special education); legislation requiring nonbiased assessment; differing conceptions and empirical results regarding bias in tests; prerequisites to implementation of nonbiased assessment procedures (including multidisciplinary contributions, outcomes criteria, and placement options and effective programs); and specific changes in assessment procedures (such as minimizing the risk of classfication). Among suggestions are that efforts be continued to insure that assessment procedures result in positive benefits for individuals, that the idea of multifactored assessment be implemented, and that nonbiased assessment be viewed as a process rather than a set of instruments. Appended materials include the Guadalupe court decision, P.L. 94-142 (Education for All Handicapped Children Act) rules and regulations regarding assessment and placement procedures, the Northeast Regional Resource Center outline of the nonbiased assessment process, and a rough draft of special education assessment procedures. (SBH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157240

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NONBIASED ASSESSMENT and SCHOOL PSYCHOLOGY

Daniel J. Reschly
Department of Psychology
Iowa State University
Ames, Iowa 50011

August, 1978

EC 112217

J 00445

State of Iowa
DEPARTMENT OF PUBLIC INSTRUCTION
Grimes State Office Building
Des Moines, Iowa 50319

STATE BOARD OF PUBLIC INSTRUCTION

Jolly Ann Davidson, President, Clarinda
John E. ven der Linden, Vice-President, Sibley
Cornelius Bodine, Jr., Sioux City
William N. Cropp, Des Moines
Karen K. Goodenow, Wall Lake
Virginia Harper, Fort Madison
T. J. Heronimus Grundy Center
Robert G. Koons, Clinton
Susan M. Wilson, Waterloo

ADMINISTRATION

Robert D. Benton, State Superintendent and Executive Officer
of the State Board of Public Instruction
David H. Bechtel, Administrative Assistant
James E. Mitchell, Deputy State Superintendent

Pupil Personnel Services Branch

Drexel D. Lange, Associate Superintendet
J. Frank Vance, Director of Special Education
Jeff Grimes, Consultant, School Psychological Services

ACKNOWLEDGMENTS

A number of persons contributed in significant ways to the preparation of this paper. Jeff Grimes of the Iowa Department of Public Instruction originally suggested the development of the paper, and provided support and encouragement through all phases of the project. The supervisors of psychological services in the fifteen area education agencies provided constructive criticism of the initial outline and draft of the paper. A number of persons devoted considerable time and effort toward evaluation of the organization, conceptual points, and writing style. I am particularly appreciative of the efforts of Gary Ross-Reynolds of Mason City, Iowa; Jim Tucker of the Texas Regional Resource Center; George Hynd of Northern Arizona University; Richard Kicklighter of the Georgia State Department; Kathryn Gerken of the University of Iowa; and William Wynveen of Blue Island, Illinois.

D.J.R.

TABLE OF CONTENTS

OVERVIEW	1
SPECIAL EDUCATION LITIGATION	4
Hobson vs Hansen	4
Diana vs California	5
Guadalupe vs Tempe	8
Larry P. vs Riles	9
Summary	11
UNDERLYING ISSUES IN COURT CASES	13
Nature vs Nurture	13
Meaning of IQ Test Results	15
Labeling Effects	15
Effectiveness of Special Education	18
Disproportionate Numbers Equated with Bias	19
Summary	22
LEGISLATION REQUIRING NONBIASED ASSESSMENT	22
BIAS IN TESTS: DIFFERING CONCEPTIONS AND EMPIRICAL RESULTS	25
Content Bias	26
Atmosphere Bias	29
Bias in Use: Prediction	31
Bias in Use: Social Consequences	32
Summary	33
PREREQUISITES TO IMPLEMENTATION OF NONBIASED ASSESSMENT PROCEDURES	34
Procedural Safeguards	34
Multidisciplinary Contributions	35

Outcomes Criterion 36

Placement Options and Effective Programs 38

Summary 39

SPECIFIC CHANGES IN ASSESSMENT PROCEDURES 40

Problems in Classification of Children 40

Classification: Minimizing the Risk 41

Classification: Maximizing the Benefits 44

Summary 45

Multifactored Assessment 46

Summary 49

Multifactored Assessment: Primary Language 49

Multifactored Assessment: Adaptive Behavior 51

Summary 55

Multifactored Assessment: Sociocultural Background 56

Summary 62

NONBIASED ASSESSMENT: SOME TENTATIVE CONCLUSIONS 63

REFERENCES 66

APPENDICES 72

APPENDIX A. GUADALUPE COURT DECISION 73

APPENDIX B. 94-142 RULES AND REGULATIONS REGARDING ASSESSMENT AND PLACEMENT PROCEDURES 81

APPENDIX C. NORTHEAST REGIONAL RESOURCE CENTER OUTLINE OF NONBIASED ASSESSMENT PROCESS 82

APPENDIX D. ROUGH DRAFT: IOWA DPI COMMITTEE. SPECIAL EDUCATION ASSESSMENT PROCEDURES 91

OVERVIEW

When in danger, When in doubt
Run in Circles
Yell and Shout

The issues of bias in tests and bias in assessment have provoked high frequency behaviors of the type suggested in this anonymous quote. Much heat has been generated through the yelling and shouting, but relatively little light. Illumination of improved practices in psychology and education, especially procedures which would expand opportunities and improve competencies for children, has been conspicuously absent in most of the discussions.

Perhaps the main difficulty stems from a focus on the wrong problems and the wrong questions in the discussions of nonbiased assessment. The major concern has been with the assessment of minorities, particularly questions related to whether specific tests are biased or unfair when used with Black, Latino, or Native American children. The issues related to assessment of children from minority backgrounds are legitimate and important. However, a more important issue is insuring educational experiences which maximize competencies and opportunities. The major problem then in nonbiased assessment is insuring usefulness and fairness of assessment and interventions for all persons. The focus on usefulness and fairness of both assessment and interventions provides a broader perspective on the problem of nonbiased assessment, and directs our attention beyond the typical and narrow questions of whether this item or that item, this test or that test is biased and unfair. Consideration of outcomes of assessment

can improve the assessment practices used with all children as well as reduce the alleged bias in current assessment practices used with minority children.

The purposes of this paper are to discuss the various problems related to the current concerns with nonbiased assessment. Nonbiased assessment is not simply a debate within academic settings. Legislation has been enacted, judicial inquiry has occurred and continues, and various directives from federal and state agencies have resulted. The issues which led to these events are discussed and implicit assumptions clarified. Most important, specific recommendations are made which reflect an attempt to provide a guide to school psychologists for our efforts to insure quality in assessment for all children including nonbiased assessment with minority persons.

Concerns about bias in assessment did not originate in school psychology, and can perhaps best be understood within the context of the historic process of removing vestiges of race, class, and ethnic discrimination. Similar issues of bias in assessment that will be discussed within the context of school psychology are present in the broader contexts of public education, employment settings, and professional school selection. Indeed, recent events such as resolutions suggesting the outright elimination of all standardized tests (Rudman, 1977) and the Bakke court case (Bakke vs California) demonstrate the widespread contemporary concerns over bias and usefulness. Furthermore, concerns over bias in tests are not a recent phenomenon. Issues similar to those debated today were the topic of heated discussions in the 1920s, discussions which, like those today, appeared in both the professional and popular literature (Block and Dworkin, 1976).

Perhaps the most important issue, often ignored in current and previous debates on bias in tests, is the net effect of standardized tests in the process of eliminating discrimination and expanding opportunities. There was a close parallel in the 1920s to the current discussions of the hereditary potential for intellectual development of specific racial or ethnic groups. Only in the 1920s, the primary concern was over the "innate abilities" of southern and eastern Europeans and persons of Jewish ancestry (Kamin, 1974). Tests were also used to document the hereditary inferiority of these groups. However, it is highly likely that the same kinds of tests, used earlier as proof of inferiority, served as instruments of social mobility for many individuals of Jewish and eastern or southern European ancestry. The net effect of standardized tests on eliminating discrimination against various groups has probably been positive for many groups, and more importantly, suggests a different set of criteria for us to consider in current discussions of test bias. The most important criteria in judging the effects of standardized tests are usefulness and fairness, especially in terms of outcomes for individuals. The debate over alleged bias in this test or that test probably cannot be resolved. More importantly, whether or not cultural bias exists in the test may be largely irrelevant to the more important question of insuring more effective educational outcomes and expanded opportunities for children.

Although the concerns about bias in tests are not new, and did not arise only within school psychology, school psychology and intelligence tests have been among the most active areas for expressions of concern. Perhaps the most dramatic expressions of concern have occurred in the form of litigation over the past ten years. The litigation bearing upon bias in tests was concerned with special education placement, principally the placement of

minority group students in classes for the mildly retarded. The court decisions, more accurately consent decrees, have had a significant impact on recent legislation and the practice of school psychology.

SPECIAL EDUCATION LITIGATION

Two types of court cases have been extremely influential upon recent events in special education. One type of case was initiated by parents of handicapped children, usually moderately or severely retarded, against public schools. Parents contended that constitutional guarantees existed to protect children from discrimination solely because of handicapping conditions. The fact that many school districts did not admit children with more severe handicaps was discriminatory according to the parents. In a landmark decision (PARC, 1971) a Pennsylvania federal district court upheld the contentions of parents and ordered the state to provide free and appropriate educational services for handicapped children.

The second type of case was concerned with the overrepresentation of minority students in special education programs for the mildly retarded. A number of cases have been before the courts on the placement issue. The net effect of these cases was to raise serious questions about traditional assessment practices, institute a variety of protections for parents and students, and establish guidelines for assessment of minority students. The effects and implications of these cases are sufficiently far reaching to warrant closer examination.

Hobson vs Hansen. The Hobson vs Hansen case (1967), although not directly related to special education placement, considered issues and established a precedent which appeared in subsequent cases. Hobson was the first

case in which the courts considered technical issues related directly to psychological testing. The basic issue was the constitutionality of the tracking system used in the Washington, D.C. public schools. A group ability test (the Otis-Lennon) was one of the bases for constituting a five level ability tracking system. Although the test was not the only source of information used to assign students to an ability level, it was apparently the most identifiable component at least in the deliberations of the court. The test per se was not ruled unconstitutional, but rather the use of the test, on the basis of evidence that Black students were overrepresented in lower tracks and underrepresented at higher (gifted) levels. The court reasoned that since the outcome of the test was discriminatory, the test itself was biased or unfair to Black students.

In retrospect, it is obvious that the reasoning of the court in the Hobson vs Hansen case was clearly relevant to issues in special education classification where poor and minority students were also overrepresented. The overrepresentation of minorities in classes for the mildly retarded was certainly well known in the 1960s, but was regarded widely in the literature and in university classes as the "natural" outcome of "cultural-familial" factors. School psychologists, the author included, should probably have recognized the obvious parallels of this case to special education practices, and undertaken appropriate corrective action or collected data to defend overrepresentation. In point of fact most of us expressed curiosity over the decision and ignored the implications.

Diana vs State of California (1971). The Diana case was the first court action related directly to overrepresentation of minorities in special education. Diana was a class action suit filed on behalf of nine bilingual children in Monterey County, California who were placed in special education

classes for the mildly retarded. The plaintiffs contended that inappropriate intelligence tests were used to certify eligibility for the programs. The tests were regarded as inappropriate due to emphases on verbal facility and middle class values and symbols. Further the tests were administered in English although the children enrolled in the classes were from homes in which Spanish was the primary language (and possibly the primary language of the children). The plaintiffs also contended that through the use of the tests the children were denied the Fourteenth Amendment rights to equal protection of the laws, specifically the right to equal educational opportunities. Implicit in this contention was the assumption of ineffectiveness of the special education program and the deleterious effects of the label educable mentally retarded. The crucial, and apparently most persuasive, evidence was the percentages of Spanish surnamed children in the district in comparison to the percentage of Spanish surnamed children enrolled in special education classes for the mildly retarded. In Monterey County, Spanish surnamed students constituted 18.5% of the total school enrollment, but one-third of the enrollment in classes for the mildly retarded was Spanish surnamed.

The suit was settled out of court through negotiations between the plaintiffs and representatives of the school district and the California State Department of Education. A consent decree was then issued by the court which had far reaching effects upon assessment procedures in California and implications for assessment elsewhere. It might be noted that a consent decree does not have quite the same legal status as a judicial opinion. The former apparently does not establish a case law precedent that is applicable to situations beyond those cited in the consent decree.

The consent decree in Diana required the following:

1) Assessment of primary language competence prior to administration of other assessment procedures. If the child's primary language competence was determined to be Spanish, subsequent assessment procedures had to be administered in Spanish.

2) Unfair portions of current tests such as knowledge of English word meanings were to be deleted, and greater emphasis was to be placed on the results of nonverbal or performance measures.

3) All bilingual children enrolled currently in special classes for the mildly retarded were to be reevaluated within a short time period using procedures consistent with points 1 and 2 above.

4) School districts and the State Department were required to develop services to assist those children who were returned from special education to regular classes as a result of the reevaluations (See Yoshida, MacMillan, and Meyers, 1976 for a description of the services and an evaluation of the effects of "decertification.")

5) The California State Department was required to develop and standardize a more appropriate test for Latino youth. (Author's note: Apparently this project was never initiated.)

6) Finally, the consent decree included a rather strongly worded warning to districts that disproportionate numbers of any ethnic or racial group placed in special education programs must be explained and justified.

Two direct and immediate outcomes of the Diana case can be identified. First the court decision resulted in the immediate return of several thousand Spanish surnamed children to regular classrooms. Although the transition services mandated by the courts for these children were apparently haphazard, the social and academic adjustment of many of the students was fairly positive (Yoshida, MacMillan, and Meyers, 1976). The fact that many Spanish surnamed students, previously classified as Educable Mentally Retarded, were "successful" in regular classroom programs raises a number of interesting

questions. Were the original diagnoses accurate? Did the children simply become more "Anglicized" with increasing age, develop more competence with English, and thereby adjust successfully to the regular classrooms where they apparently had failed at an earlier age? Did the special class program assist many children in overcoming adjustment or academic deficiencies which led to the original referral? Answers to these questions are not available. However these data, at a very minimum, illustrate the need for periodic and thorough review of placement decisions, and perhaps demonstrate the wisdom of routinely returning special class students to regular classrooms on a trial basis (Hewett and Forness, 1974).

The second direct outcome of the Diana case was an Office of Civil Rights (OCR) Memorandum issued in 1972 (Oakland, 1977) which specified certain procedures to guide the assessment of minority students. This memorandum was particularly concerned with the possible relationship of overrepresentation of minority students in special education to the broader issue of segregation in the public schools. A copy of this and other memoranda from OCR are included in Oakland (1977).

Guadalupe vs Tempe Elementary District (1972). The Guadalupe case was nearly identical to the Diana case in terms of issues (overrepresentation) and method of resolution (consent decree). Guadalupe is important to this discussion in that the consent decree went even further in specifying assessment procedures. In addition to the same requirements concerning assessment included in Diana, Guadalupe required the following: (see Appendix for a copy of the consent decree).

1) No child shall be placed in programs for the mildly retarded unless the intelligence test score is two or more standard deviation below the mean. Note - the Arizona educational definition of mental retardation at that time was IQ of less than 75 and the AAMD criterion prior to 1973 was one standard deviation or more below the mean.

2) Intelligence test results shall not be the exclusive or the primary basis for classifying children as mentally retarded. Note - This same statement has appeared frequently in legislation and rules and regulations over the past five years.

3) If the child's primary language competence was determined to be in some language other than English, classification decisions were to be based upon non-verbal or performance types of measures.

4) Assessment of adaptive behavior through, but not limited to, an interview with the parents or guardian in the child's home.

The Guadalupe case also required several other procedures which were later incorporated in federal legislation such as informed consent for evaluation and placement, due process, integration of programs for handicapped and normal students, and accountability of school districts in terms of data on "prior effectiveness of special education programs." (Guadalupe Court Decision, p. 6)

Larry P. vs Riles (1972, 1974, and pending). The Larry P. case is also a class action suit related to the basic issue of overrepresentation of minority students in programs for the mildly retarded. Larry P. was filed on behalf of seven Black children (and all others like them) placed in programs for the mildly retarded. The case was filed originally in November, 1971; an injunction was issued by the Federal District Court for Northern California in June, 1972; an expanded injunction was issued in 1974, and the case is currently (finally) before the court.

The preliminary injunction in Larry P. restrained the San Francisco School District from

> "placing black students in classes for the educable mentally retarded on the basis of criteria which place primary reliance on the results of IQ tests as they are currently administered, if (emphasis added) the consequence of use of such criteria is racial imbalance in the composition of such classes" (Larry P. vs Riles Court Injunction, 1972).

In 1974 the plaintiffs requested the court to expand the original injunction to include all school districts in the State of California. The injunction was expanded in 1974 which prompted the California State Board of Education to forbid the use of individual intelligence tests for all students in California schools IF the outcome of such tests was a classification decision of mental retardation (News Release, California State Board of Education, January 15, 1975). Intelligence tests have not been banned generally by the courts in California as is commonly believed, and use of intelligence tests in California is still permissable as long as a decision of mental retardation is not under consideration.

The Bay Area Association of Black Psychologists was instrumental in developing and pressing the case against the San Francisco School District and the State of California. The defense in the case has been assumed by the Attorney General's Office in California with the assistance of the State Department of Education. Both sides have or are presenting massive amounts of evidence in terms of documents or testimony from expert witnesses from around the country. The Larry P. case is potentially much more far reaching than any of the previous court suits. The case was filed on the basis of constitutional law in a federal district court, and the case will apparently be decided by judicial opinion rather than a consent decree negotiated out of court. Finally, both sides in the case are reportedly committed to appeal of the decision, possibly to the Supreme Court. Although the Larry P. case may seem remote and irrelevant to school psychologists outside of California, the Larry P. decision may have a decisive influence on the nature of assessment throughout the country.

The basic issues in Larry P. include the same concerns addressed in Guadalupe and Diana with certain crucial additions. In both Diana and

Guadalupe the plaintiffs requested revision and reform of current assessment practices. In Larry P. the plaintiffs have requested elimination of all standardized tests including, of course, individual intelligence tests. The plaintiffs have argued that the resolution of previous cases such as Diana did not go far enough. To substantiate this assertion the plaintiffs cited data in 1974 which indicated large overrepresentation of ethnic or racial minorities in special education programs in California despite the Diana decision. In both 1972 and 1974 when injunctions were issued by the court, the plaintiffs argued that Black children were overrepresented in programs because of inherent biases in the tests. Further, the plaintiffs contended the programs were ineffective, the labels attached to children were stigmatizing and humiliating, and the overrepresentation constituted an abridgment of the Fourteenth Amendment guarantee of equal protection. Although a number of issues in addition to cultural bias in tests were implicit in the contentions of the plaintiffs, the court focused on the issue of test bias in both of the preliminary injunctions. Apparently, other issues such as effectiveness of programs, alleged labeling effects, preplacement adjustment of students in programs and alternative assessment procedures are now being considered in the evidence presented by the plaintiffs and defense (Personal Communications, Jane Mercer, Richard Russo, and Jerome Sattler, November-December, 1977). However, the problem of test bias remains a very important, perhaps crucial issue in Larry P.

Summary on Litigation. The litigation in special education has clearly been a crucial influence on state and federal legislation, and on professional practices in school psychology. The PARC decision along with other similar cases (Mills, etc.) was the primary impetus for the enactment of mandatory special education legislation by the states and the federal Education for All

Handicapped Children Act of 1975 (PL 94-142). On one hand the litigation has led to a dramatic expansion of special education services. On the other hand litigation has raised serious doubts about the effectiveness and fairness of conventional special education procedures. All of the basic provisions of PL 94-142 appeared earlier in one form or another in litigation. Although other factors were certainly important in influencing the nature and enactment of 94-142, litigation must be recognized as one of the most important factors.

It is also important to understand the judicial mechanism as a means of resolving professional or scientific issues. The fundamental purposes of the courts are somewhat different from the aims of science or professions. The legal system is primarily concerned with justice where science is concerned primarily with truth. The issues addressed by the courts in the special education placement litigation are at best ambiguous. These issues, e.g. labeling effects, bias in tests and effectiveness of special education programs are all sources of intense debate within the scientific and professional communities. None of these issues can be resolved "beyond the shadow of doubt" with the currently available empirical evidence. All of the issues mentioned above, however, are related to rights and opportunities of persons which are of course legitimate areas of judicial inquiry. The problem is that the courts, usually reluctantly, must ultimately resolve the issues on the basis of very ambiguous, sometimes technical evidence. The courts, probably reluctantly, resolve the disputes in "absolute" language such as "shall", or "must" when in fact the evidence is at best at the level of "might" or "should".

One very important feature of the litigation to date, especially Larry P., is the question of burden of proof. In the Larry P. case the court shifted the burden of proof to the defendents on the basis of the evidence on over-

representation. The court viewed the fact of overrepresentation as "inherently suspicious" from the broader context of the legality of segregation. The burden of proof is crucial in that it is probably impossible to prove or disprove unequivocally whether or not standardized tests are biased on the basis of the available evidence. The fact that the defendants have been responsible for proving their case has been a distinct disadvantage. However, as will be discussed in the next section, many other issues are implicit in the special education litigation on placement. These issues are probably more important in terms of the rights and opportunities of children than the narrow issue of test bias.

UNDERLYING ISSUES IN COURT CASES

A number of unresolved issues and implicit assumptions are apparent from close examination of the previous and current litigation over special education placement. These issues involve fundamental questions about the nature of intelligence, the direct and indirect effects of classification and educational services for the mildly retarded, and the outcomes of intelligence tests for culturally different persons. The court decisions have appeared to ignore many of these issues. Understanding these issues is crucial to the development of fairer and more effective assessment procedures and educational interventions.

Nature vs Nurture. The nature vs nurture debate over intellectual differences is an old controversy which even precedes the development of IQ tests (Galton, 1892). The issue is far from new, the general outline of the debate has not changed, and the issue is no closer to resolution now than it was in the 1800s. However, the debate over the past ten years has been

intensified for a variety of reasons. Arthur Jensen's (1969) highly controversial evaluation of and explanation for the disappointing effects of Head Start is a crucial component of the current attacks on intelligence tests. Jensen has published widely on the alleged genetic inferiority of Blacks. Although Jensen's conclusions are usually stated in the tentative and conditional style favored and required by the scholarly community, the conclusions of several other "scholars" have been nothing short of inflammatory. For example, the following statement which appeared a few years ago in a highly regarded academic journal was based largely on Jensen's conclusions.

> "Nature has color coded groups of individuals so that statistically reliable predictions of their adaptability to intellectually rewarding and effective lives can easily be made and profitably be used by the pragmatic man in the street." (Schockley, 1971, p. 377) (emphasis added)

The statement is outrageous and untrue. Therefore, it is not hard to understand the sense of outrage particularly among Blacks that arises from this and similar statements. In view of such statements, I am periodically inclined toward eliminating all standardized tests, or for that matter, any other measure which is cited to justify racist ideology.

A complete review of all of the data on the nature-nurture issue is far beyond the scope of this paper (See Brody and Brody, 1976; Loehlin, Lindzey, and Spuhler, 1975; and Samuda, 1975). Two points must be emphasized. First, elimination of standardized tests would not settle the nature-nurture controversy, and would, in all likelihood, reduce the chances of overcoming existing barriers to the full participation of all persons in the economic and social order. Secondly, the professional personnel responsible for using the tests have an ethical and moral obligation to insure that test results are interpreted accurately and test use leads to expanded, not diminished, opportunities and competencies for children.

Meaning of IQ Test Results. Perhaps the most immediate challenge for professional personnel who use standardized tests is to once again attempt to clarify the meaning of IQ test results. Alfred Binet was the first to warn of the misuse and misunderstanding of the meaning of IQ test results. Others have issued similar warnings over the past seventy years. Certain myths about IQ are all too prevalent for us to ignore. Many consumers continue to believe that IQ is fixed, unitary, and predetermined. Much testimony in the court cases including Larry P. (APA Monitor, 1977) was directed at disproving those myths. This testimony has a "Straw man" quality. It is not hard to substantiate charges which confirm the obvious. Unfortunately, we have not done enough to eliminate the misconceptions about IQ. I have suggested that we might develop kind of a "Surgeon General's Warning" about IQ which might be printed on every test protocol, every test report, and placed in every file where IQ test results are included. The following statement is not perfect, and could be improved with the help of other persons.

> "IQ tests measure only a portion of the competencies involved with human intelligence. The IQ results are best seen as predicting performance in school, and reflecting the degree to which children have mastered middle class cultural symbols and values. This is useful information, but it is also limited. Further cautions - IQ tests do not measure innate - genetic capacity and the scores are not fixed. Some persons do exhibit significant increases or decreases in their measured IQ."

Unfortunately, one of the issues in the litigation has been whether or not measured intelligence is unitary, fixed, and predetermined. The testimony which once again proves these assertions to be myths may be important to the resolution of the Larry P. case.

Labeling Effects: Meaning of Mild Mental Retardation. A closely related set of implicit assumptions in the court cases involves the meaning of mild mental retardation, labeling effects, and efficacy of special classes for the

mildly retarded. The courts have apparently been convinced that the label Educable Mentally Retarded is humiliating, stigmatizing, and solely responsible for a negative self-fulfilling prophecy. Further, the courts have examined the term mental retardation, and apparently have understood it to imply permanent, global incompetence with a biological etiology (APA Monitor, 1977). The available factual evidence related to both of these assumptions reveals a quite different state of affairs.

The alleged effects of labels have dominated much of the discussion of the outcomes of special education programs over the past ten years. Four basic questions seem to be especially important in this discussion. First, do labels create expectancies? The evidence generally confirms this assertion, however, the outcomes of these studies may be an artifact of the methodology used. In studies where subjects (college students or teachers) are given only the label and/or no or only brief exposure to the child bearing the label, a rather large expectancy effect is typically reported (e.g., Foster and Ysseldyke, 1976). The studies essentially involve telling the subject that Johnny is retarded, then asking the subjects to rate Johnny's likely or briefly observed performance on an academic task. In other studies using the same basic methodology, but providing lengthy exposure to the "labeled child", the expectancy effect is not observed or diminishes over time (Yoshida and Meyers, 1975 or Lamprecht, 1977). Do labels create expectancies? The evidence is contradictory from experimental studies, and few studies have been conducted in natural situations.

The relationship of expectancies to a self-fulfilling prophecy (Rosenthal and Jacobson, 1968) i.e., the child performs markedly better or poorer due to expectancies of significant others, is even less clear. This basic idea of the self-fulfilling prophecy is widely cited as fact, and apparently is

consistent with certain philosophical or political orientations. However, the data simply do not support the assertion (MacMillan, Jones, Aloia, 1974; Humphreys and Stubbs, 1977).

A related, and usually unexamined, issue in the litigation is the kinds of informal labels used in the public schools and the "prelabeling" experiences of children classified as mildly retarded in the schools. From reading the testimony in these cases one is perhaps led to believe that no one ever saw those children as academic or behavioral problems prior to being labeled by a psychologist. The erroneous impression that school psychologists and IQ tests were the first and most important step in classifying school children as mildly retarded has been refuted (Meyers, Sundstrom, and Yoshida, 1974), but this misconception is prominent in the litigation. The fact is that teacher referral is the most important step in the process whereby children are classified as mildly retarded. Many children referred as suspected cases of mild retardation are not diagnosed as such by psychologists (Ashurst and Meyers, 1973), and many children who would fail the IQ criterion are never referred (Mercer, 1973). The behaviors which lead to referral are as important if not more important than the IQ tests per se. The courts have generally ignored these factors.

Evidence on the final questions of the meaning of the diagnosis of mild mental retardation and the reactions of persons so labeled is much clearer. Most persons misinterpret the meaning of mild retardation. Mild retardation according to the 1961 and 1973 American Association on Mental Deficiency (AAMD) Manual on Terminology and Classification refers to the current behavior of the person (i.e., is not permanent), is developmental in nature (i.e., different criteria are used in assessment at different ages), and the etiology is not specified, (i.e., may or may not be associated with biological

pathology). In fact, the AAMD system has implicitly suggested environmental factors such as cultural-familial or sociocultural as the primary causes of mild retardation. The association of these factors with the effects of poverty has been recognized for a long time, and continues to be the source of much speculation and research (Haywood, 1970; Heber, et al., 1972). Most important to the court deliberations, is the well known fact that the vast majority of the mildly retarded are deficient only in a fairly narrow range of competencies (those associated with formal education), and most become self-supporting and independent adults (Baller, Charles, and Miller, 1967). The widespread misconceptions about the meaning of mild retardation suggest the need for a revision in the classification system as well as a concerted effort to clarify the meaning of the term mild mental retardation.

Data on the reactions of persons who have been or who are classified as mildly retarded are much clearer. Persons who bear this label regard it as stigmatizing, humiliating, inaccurate, and unfair (MacMillan et al., 1974; Edgerton, 1967). The persons who bear the label have the same misconceptions about the meaning of mild retardation cited earlier, i.e., comprehensive, permanent incompetence of a biological origin. The persons labeled as mildly retarded often cite their present level of adjustment as irrefutable evidence of the inappropriateness of the label. The problem continues to be the misconception about the meaning of the classification, an issue which will be discussed again in a later section.

Effectiveness of Special Education Interventions. Perhaps the most important allegation by plaintiffs, implicitly accepted by the courts, was the presumed ineffectiveness of special classes for the mildly retarded. The effectiveness of these programs has of course been questioned (Dunn, 1968) and debated (Kolstoe, 1976). The defendants in the cases (States and local

districts) have not to date attempted to defend the efficacy of the educational programs in benefiting the children classified as mildly retarded (the defense in Larry P. has not yet been presented). Testimony from parents and others has focused on the ineffectiveness of the programs. Since the defendents have not even attempted to rebut this testimony, the courts have assumed the programs were inferior to regular classrooms. The court decisions have focused on IQ tests and the overrepresentation of minority children. However, if the educational programs were as ineffective as described by the parents, then the programs as such were a denial of the constitutional rights of any child regardless of racial or ethnic status!

Furthermore, the litigation may reflect "worst case" situations (MacMillan, 1977). Special education programs for the mildly retarded generally, (and the psychological services associated with these programs) may not have been represented accurately by the situations brought to litigation. MacMillan has pointed to a wide variety of inappropriate services involved with the Diana case (e.g. students in classes for the mildly retarded were used to perform janitorial work, wash buses, and even dismissed from school to work as farm laborers - these activities were apparently not part of a work study or career education program). It is likely that the states and districts did not defend their educational or psychological services because the services were indefensible! The degree to which these services are typical of most states or districts is unknown.

Disproportionate Numbers Equated with Bias. Evidence on the overrepresentation of minority students in classes for the mildly retarded has been a major component in all of the court decisions cited earlier and still represents the major emphasis of the federal Office of Civil Rights. Overrepresentation of minorities has been regarded as inherently suspicious by

the courts, and has resulted in shifting the burden of proof from plaintiffs to defendants in the Larry P. case.

Two facets of the overrepresentation data bear closer analysis. First, the overrepresentation data are sometimes misinterpreted and/or exaggerated in attacks on the uses of intelligence tests with minorities. The percentages can be very misleading if not understood properly. For example, in the Larry P. case Black students constituted 28.5% of the total district enrollment, but nearly 66% of the enrollment in special education classes for the mildly retarded. These data have sometimes been understood to mean that two-thirds of all Black students were diagnosed as mentally retarded through the use of intelligence tests. In fact, as Table 1 illustrates, a much smaller percentage of Black students were actually diagnosed as mildly retarded.

Table 1

APPLICATION OF OVERREPRESENTATION DATA TO A HYPOTHETICAL POPULATION OF STUDENTS

28.5% of the student enrollment is minority
66% of the enrollment in programs for the mildly retarded is minority
2% of the total student population is enrolled in special education programs for the mildly retarded
Then apply these data to a hypothetical population of 40,000 students
28.5% or 11,400 students are minority
2% or 800 students are in programs for the mildly retarded
66% or 528 minority students are in programs for the mildly retarded
528 ÷ 11,400 or 4.6% of the minority students are classified as mildly retarded

The only assumption made in the illustration in Table 1 is the assumption that 2% of the total student population is classified as mildly retarded. The assumption of 2% is an estimate based upon actual special education enrollment data from a variety of sources. The actual enrollment for a specific district may be slightly higher or lower. The main point is that even with the unusually large degree of overrepresentation involved with the Larry P. case, only a relatively small percentage of minority students were actually classified as mildly retarded. These data certainly do not support the assertion that the primary function of IQ tests is to label minority students as "uneducable and retarded."

Curiously, persons with other equally obvious demographic characteristics are also overrepresented in special education programs. For example, the ratio of males to females in programs for children with mild retardation or learning disabilities is at least 2:1, and probably much higher. More important to our present discussion is the overrepresentation of children from economically poor homes in programs for the mildly retarded. If we divided a student population into two groups simply on the basis of median family income, and then analyzed the income characteristics of students in programs for the mildly retarded, I would wager that even greater disproportionality would result. The fact is that very few children from middle and upper class homes obtain intelligence test scores within the mildly retarded range which of course is one of the criteria for the diagnosis of mild mental retardation. This relationship, i.e., the association of poverty with mild mental retardation along with the income characteristics of certain minority populations, raises an intriguing question. Are minorities overrepresented in programs for the mildly retarded because of minority status or because of socioeconomic status?

The purpose of the questions raised above is not simply to defend the overrepresentation of minorities in special education. As mentioned earlier, if the programs for the mildly retarded were (are?) as bad as alleged, then placement in such programs is inappropriate for any child regardless of racial or ethnic status. However, these questions may direct our attention to more relevant issues, e.g., the usefulness and fairness of assessment procedures generally, rather than the narrow preoccupation with test bias which unfortunately has been the focus of the litigation.

Summary. A number of assumptions were apparently made in the litigation concerned with special education placement. These assumptions, although largely unexamined in testimony, were apparently crucial to the court decisions. The criticisms of these assumptions contained in this paper are not meant to justify what has been done, or what has been common practice in the schools. These criticisms will hopefully assist us in focusing our attention on the most important issues, issues related to improving the usefulness and fairness of assessment data for all students. Nonetheless, the court decisions have focused on bias in tests and have been a significant influence on recent legislation.

LEGISLATION REQUIRING NONBIASED ASSESSMENT

Two recent Federal laws provide nearly identical guidelines for the assessment and placement of handicapped children. Section 504 of the Rehabilitation Act of 1973 (Rules and Regulations, May 4, 1977) and PL 94-142, The Education for All Handicapped Children Act of 1975, (Rules and Regulations, 1977) reflect the clear influence of the special education placement liti-

gation discussed earlier. Most pertinent to this paper is the following requirement from 94-142.

> "Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory."

This requirement, commonly referred to as nonbiased assessment, is one of the most important features of the legislation, but potentially, one of the weakest since no clear definition and only limited discussion of the meaning of bias in assessment appeared in the rules and regulations. In view of the differing interpretations and contradictory evidence on bias in tests or assessment, the practical effects of the requirement are unpredictable.

Perhaps in response to the potential confusion, the Bureau of Education for the Handicapped (BEH) awarded a large contract to the Coordinating Office for Regional Resource Centers (CORRC) to explore the meaning of bias in assessment. The CORRC project has issued three reports on bias in assessment. Volume II of this series entitled "With Bias toward None" (Also available as a book, Oakland, 1977) is the most substantive and pertinent of the reports. The CORRC projects have not resulted in clearly stated definitions of bias or specific guidelines for eliminating bias in assessment. This is not surprising in view of the current state of the art. Both the CORRC reports and the federal rules and regulations have apparently concluded that although bias cannot be defined unequivocally, evaluation procedures are likely to be less biased if procedural safeguards are followed and a broad variety of information is gathered and considered.

Copies of the requirements for evaluation and placement in the Federal laws are provided in an appendix at the end of this paper. The most important

features of these guidelines are:

1) Procedural safeguards which provide for informed consent and due process are required.

2) The assessment must be conducted in child's native language if at all possible.

3) Tests and other evaluation devices are validated for the specific purpose for which they are used, and administered by trained personnel.

4) Classification and placement decisions are not based on a single source of information (such as IQ) and areas of specific educational need are identified in the evaluation process.

5) Inferences about aptitude or achievement are not made from evaluation procedures which reflect the child's impaired sensory, manual, or speaking skills. (Author's translation: inferences about the abilities of a blind child cannot be made on the basis of responses to a performance test).

6) Assessment must be conducted in a broad variety of areas and placement procedures shall draw upon information from aptitude and achievement tests, teacher recommendations, physical conditions, social or cultural background, and adaptive behavior. (emphasis added). Further, information from the above sources must be documented and carefully considered.

7) Decisions are made by a multidisciplinary team with participation of parents.

8) Placement options are selected according to the principle of least restrictive alternative and an individualized educational plan is developed.

9) The educational program is reviewed annually and a comprehensive reevaluation which meets the requirements stated above is conducted at least every three years.

It might be noted that the nonbiased assessment requirement appears to be concerned primarily with two populations of children. First, there is the obvious concern about the kind of assessment conducted with the culturally different individual which has been the focus of this paper. Secondly, there is concern (see No. 5 above) with the assessment of children with sensory,

manual, or speaking impairments which has not been discussed here (see Gerken, in press). These legislative guidelines are nearly identical to consent decrees issued by the courts in the early 1970s (see Guadalupe consent decree in the appendix). The guidelines are stated in general terms with few precise or specific suggestions for practitioners. For example, consider the question of sociocultural background which, according to the legislation must be considered, and the consideration documented, in making placement decisions. There is no specification of how sociocultural background is to be assessed or how it might be taken into account by the multidisciplinary team. These issues which are considered in a later section, will undoubtedly be the subject of much discussion and research in the future.

BIAS IN TESTS: DIFFERING CONCEPTIONS AND EMPIRICAL RESULTS

There have been many efforts to define bias in tests and assessment (e.g., Novick and Peterson, 1976; Oakland and Matuszek, 1975; Hunter and Schmidt, 1976; and Reschly, in press) but consensus on theoretical or research criteria and agreement on practical implications have not been achieved. Analyses of bias in specific tests have ranged from speculative judgments about specific items to sophisticated statistical examinations of test results and prediction systems. The conclusions of these efforts are largely contradictory. Analyses using subjective judgments of bias usually lead to identification of many examples or sources of bias in current tests. Analyses of data from various groups usually result in conclusions of little or no bias in current tests. Throughout the discussions of test bias conclusions are confounded by confusion over or differing interpretations of the meaning

of IQ test results, an issue discussed earlier. The major outlines of these differing approaches will be reviewed in the following section.

Some agreement concerning the categories of bias that may exist in assessment procedures appears to be emerging from this very complex literature. The major elements of possible bias are increasingly organized around the concepts of content bias, atmosphere bias, and bias in use. Information on each of these kinds of bias has appeared in the literature including some research evidence.

Content Bias. Allegations of cultural bias in the items used on conventional tests has been and continues to be the most popular of the criticisms of standardized tests. In fact examination of an item from a current standardized test to support the allegation of bias in all of the items appears to be an increasingly popular indoor sport. Examples of subjective judgments of item bias are numerous (e.g., APA Monitor, 1977; Dent, 1976; or Williams, 1971). The implicit assumption is that all items on the test are biased if one or a few of the items is apparently biased. If the test is presumed to be biased on the basis of inappropriate items, then the test results are presumed to be "inaccurate" and unfair. If the items are biased, usually meaning that opportunity to learn the content of the item is not common to all environments, then the test results certainly do not reflect, and cannot be interpreted as evidence of "innate" intelligence. However, as discussed previously, the IQ test results are not direct measures of innate ability for any group.

The distinction between cultural bias and cultural loading is important to this discussion. The degree of cultural loading of an item, i.e., the likelihood of success on the item for persons with different backgrounds and experiences, varies on a continuum. At one end of the continuum are items

which could only be answered correctly by persons with highly specific backgrounds and experiences. An example might be an item which asks "Name three presidents of Iowa State University over the past century" (the present author can name only two). The item is similar to those used on many intelligence tests in terms of type of thinking required. However, only a very limited sample of persons would have an opportunity to be exposed to this information and thereby answer the item correctly. The item reflects a very high degree of cultural loading, and would be regarded by most as culturally biased. Some items on current standardized tests require similar kinds of thought and also vary in degree of cultural loading. The degree of cultural loading of an item, however, depends upon the characteristics of the persons taking the test, not the item per se.

The "person specific" nature of item bias has been illustrated well in the development of "counterbalanced" or culturally specific intelligence tests (e.g., Dove, undated, or Williams, 1975). These tests require highly specific information which is usually possessed only by persons with particular backgrounds or experiences. Knowledge of "What is a short dog", or "In "C.C. Rider, what does "C.C." stand for" or "What was a zoot-suiter" is common only to persons with very specific experiences.

In addition to subjective judgments, critics of current tests also point to the differences among various groups as evidence of item bias. The differences in average performance among various groups are attributed to item bias and/or atmosphere bias. The fact that certain groups of White Anglo Saxon Protestants (e.g. low socioeconomic status Appalachian Whites) also obtain lower scores on conventional tests is usually not mentioned by the critics, and suggests the differences are not simply due to the factors of race or ethnicity.

A solution attempted earlier, mentioned only occasionally in recent years, was the development of culture free or culture fair tests. Generally, nonverbal or performance tests have been regarded as less culturally loaded, although not all minority groups perform better on nonverbal or performance tests. Nonverbal or performance tests are now generally recognized as falling short of the goal of freedom from cultural influences, and attempts to develop culture fair verbal tests (e.g. Davis and Eels, 1951) are recognized as failures. Current thinking suggests the original concept of culture free or culture fair was probably faulty (Anastasi, 1976) in view of the usual purposes of tests, i.e., predict or evaluate performance within a cultural context.

Relatively little empirical research on or critical examination of the allegations of item bias have appeared in the literature. Subjective judgments of item bias, however, are not necessarily consistent with empirical data. The following item, "What is the thing to do if a boy (girl) much smaller than yourself starts to fight with you?" which appears on the WISC-R Comprehension subtest, has been criticized as biased against the experiences of urban Black children where it is presumably more acceptable to respond physically (incorrect answer) than verbally (correct answer). In fact, this item is <u>relatively</u> easier for Blacks than Whites. Empirical examinations of item bias for different groups have been relatively rare, but the minimal data that exist suggest caution in conclusions based only on subjective judgment. Furthermore, some of the allegations of item bias as well as the items from culturally specific tests reflect negative racial or ethnic stereotypes. The criticism of the item above implies that urban Black children are taught by parents and peers that it is acceptable to beat up smaller and younger children. I doubt very much that such attitudes or

behaviors are any more typical or acceptable among Blacks than Whites. Another example is an item from the BITCH Test (Williams, 1975). "What is Mother's Day?" (Correct answer is the day the Welfare checks arrive) suggests the incorrect stereotype that all Black children are from families supported by welfare, and is of course, unfair to the millions of Black children from intact families where the sole source of support is income from jobs held by parents.

The evidence on item bias is simply inconclusive. Test items do vary in amount of cultural loading. Items on current tests are culturally loaded to varying degrees as they must be if tests are to predict or evaluate important behaviors that occur only within a cultural context. Subjective judgments of item bias are not necessarily accurate, and revision of current tests either in the direction of greater or lesser cultural loading might have the effects of simultaneously increasing or maintaining group differences and reducing validity. The issue of validity or bias in test use will be discussed in a later section.

Atmosphere Bias. In addition to bias in content, a frequent criticism of standardized tests is that the atmosphere of the testing situation is unfair to minority children. Two general aspects of the testing environment are possible sources of unfairness: (1) The kinds of responses and nature of the effort required on the test or (2) The nature of the interaction with the examiner may be inconsistent with the child's background or experiences.

It is important to note the basic assumption of maximum performance on achievement, ability and aptitude tests. If the child cannot or does not perform as well as possible due to unique features of the testing environment, the results of the test are inaccurate reflections of the child's thinking

competencies or academic skills. In such cases comparisons of the child's performance to that of the normative sample are inappropriate.

A great amount of research has been conducted on atmosphere bias, and is well reviewed by Sattler (1970, 1973, and 1974). The interested reader is encouraged to pursue further information in those sources. The author's overall impression of this research is that:

1) Much of the research was poorly designed.

2) Some of the studies used experimental manipulations that are atypical and inconsistent with good testing practices. For example, token reinforcers provided for correct answers.

3) The results of reasonably well controlled studies in which the variables manipulated were within the range of good testing practices are contradictory. For example, the degree of warmth, amount of encouragement, time devoted to establishing rapport prior to testing, and sex or race of examiner or examinee, have been studied with mixed results.

4) Examiner expectancies for performance may influence scoring of responses on items where there is some subjectivity in evaluating responses, e.g., vocabulary subtest of the Wechsler scales.

5) When differences due to atmosphere effects are reported, the size of the differences is usually fairly small.

It should be emphasized that the generalizations above are the author's impressions of a fairly large body of knowledge. The results of this research do not necessarily generalize to all natural settings, or to the performance of all individuals. Professional personnel who administer tests to culturally different persons must be sensitive to individual variations in values, motivation, language, and cognitive style which could influence the results of the test. One of the most important roles of the examiner in individual evaluations is to establish the kind of climate that will elicit the child's maximum performance. The test results are invalid to the degree that maximum performance is not elicited.

Bias in Use. The third concept of bias in tests is concerned with how tests are used. Two different approaches to analyzing bias in use have emerged from the recent literature. One approach favored generally by leaders in test theory and psychometric research emphasizes the relationships of tests to other criteria for different groups. The second approach emphasizes the implications and outcomes of test use for individuals and groups of persons. The two approaches to bias in use will be discussed separately since they use different criteria and result in different conclusions.

Bias in Use: Prediction. Academic psychologists and test publishers have generally emphasized the technical adequacy of tests in predicting various criteria for different groups. Several definitions of test bias have been proposed to guide examinations of the degree to which tests function in the same way for persons regardless of group membership. The Cleary definition (1968) characterized test use as biased if the predictions or decisions based upon the test were different as a function of group membership. Stated simply, Cleary and most leaders in psychometric theory, see test use as unbiased if the same predictions are made for persons with the identical test scores regardless of group membership (Cleary, Humphreys, Kendrick, and Wesman, 1975). A number of variations of the Cleary definition have been proposed, most notable of which are those which stress the social utility of test use (Darlington, 1971; Novick and Peterson, 1976).

Although some of the definitions of bias in test use become very complex (Peterson and Novick, 1976), certain basic features are prerequisite to fairness in test use. Tests cannot be regarded as fair unless they predict with equal accuracy for all groups. Most pertinent to our concerns in school psychology are the questions: Do IQ tests predict academic achievement

equally well for different groups? Do IQ tests predict classroom performance equally well for different groups?

The data from a variety of studies conducted with prospective employees, prospective graduate and undergraduate students, and school age children suggest that conventional standardized tests predict the usual criteria equally well for all groups. For example, the correlations between the WISC-R and either teacher ratings of achievement or a standardized test of achievement are virtually identical for Anglo, Black, and Latino groups (Reschly and Reschly, in press). Further, the regression equations for these different groups, although not identical, are highly similar (Reschly and Sabers, in press). When differences in regression systems were found in these and other studies, the effects were overprediction for Nonanglo groups and underprediction for Anglos. The available evidence supports the conclusion that current tests predict equally well for the different groups studied thus far, and from one perspective on bias in use, conventional tests can be regarded as fair to all groups.

Bias in Use: Social Consequences. The previous definitions of test bias, although important, are inadequate in terms of the overall influence of tests upon the lives of persons. Testing does have social consequences. Tests, even those which predict accurately, have been misused to justify race, social class, and ethnic discrimination. Test results have led to reduction of opportunities for persons and have qualified persons for apparently ineffective interventions which may have been stigmatizing and humiliating. To defend tests simply on the basis of predictive accuracy is to miss entirely the points raised by recent critics of tests.

For example, Williams (1973) charged that I.Q. tests predict achievement as well for Blacks because of the intervening variable of bias. In his view,

both the predictor (IQ test) and the criterion (school achievement) are contaminated by racial bias, hence the correlation is significant and positive, but meaningless.

Jackson's (1975) response to Cleary et al.'s report of the American Psychological Association Committee on Educational Uses of Tests is even more to the point. Jackson saw the report as largely irrelevent to the concerns expressed by minorities. The report defended the technical adequacy of the tests when in fact the major concerns of Black and Chicano psychologists (Bernal, 1975) are with how tests affect the lives of persons. The fact that tests have been used to justify racist ideology, and otherwise have been misused or misinterpreted in inferences about the potential of individuals are facts acknowledged even by the authors of the APA report. Thus, to defend tests on the basis of evidence of common regression systems, or to attempt to separate the issues of technical adequacy from those of social consequences is insufficient.

The ultimate criterion that should guide our evaluations of test bias is the implications and outcomes of test use for individuals. Succintly stated, test use is fair if the results are more effective interventions leading to improved competencies and expanded opportunities for individuals. Test use is unfair if opportunities are diminished or if individuals are exposed to ineffective interventions as a result of tests.

Summary. Although numerous attempts to define and provide evidence on test bias have appeared in the literature, agreement on this crucial issue has not been achieved. The technical definitions of test bias have resulted in studies which appear to substantiate the fairness of current tests. Examinations of content bias and atmosphere bias have been largely inconclusive. The issue is far from resolved, and probably cannot be entirely

resolved with empirical data. Concerns about the social consequences of test use, especially positive or negative outcomes for individuals, must also be considered. Despite the lack of agreement on concepts of bias, recent legislation has required implementation of nonbiased assessment procedures.

PREREQUISITES TO IMPLEMENTATION OF NONBIASED ASSESSMENT PROCEDURES

A number of conditions are best seen as prerequisites to nonbiased assessment. If these conditions are not present, nonbiased assessment, or more generally, effective assessment is usually impossible. These conditions are related to procedural safeguards, multidisciplinary contributions, an outcomes criterion, and the availability of effective interventions and alternative options for serving children.

Procedural Safeguards. Perhaps the portions of the recent legislation which have been implemented most rapidly throughout the country are those which require informed consent of parents prior to evaluation and placement, and which establish procedural due process including the right to appeal decisions. It is the author's impression that nearly every school district in the nation now follows at least the guidelines for informed consent and due process. Achieving the full spirit of these requirements is much more difficult, and progress in this area is probably less consistent. Several good references exist on informed consent and due process (e.g., Abeson, Bolick, and Hass, 1975). It should be noted that both informed consent and due process have been ignored frequently by special educators and school psychologists in the past. These issues were also considered by the courts in the special education placement litigation where in several instances

parents testified that they were not even informed of the decision to place their child in a special education program. Again, we have little information on the generality of such practices. Although informed consent and due process were established as legal requirements only recently, professional ethics and standards for best practices have always emphasized the importance of good communication with and involvement of parents. It is important that we achieve the spirit and intent of informed consent and due proces. Both are best understood as processes whereby lines of communication are established and facilitated, and rights and interests protected. A good guideline for most professionals to use in evaluating their performance in carrying out the spirit of informed consent and due process is "Would you be satisfied that your rights and interests were respected and considered by the communications used if the child under consideration were your own?"

Multidisciplinary Contributions. Past standards for best professional practices and current guidelines require the involvement of a multidisciplinary team in assessment and placement of children. In particular, the guidelines require that no single source of information be used as the sole basis for placement and that a broad variety of information be gathered and considered. The source of these guidelines in the legislation may be the erroneous assumption that school psychologists and IQ tests were solely responsible for the overrepresentation of minorities in special education programs for the mildly retarded. Regardless of the source of the guideline, the intent is crucial to effective assessment and interpretation. Special education placement decisions are extremely important events in the lives of children. Since instruments (and persons) are fallible, no single person or single measure should be the basis for a special education placement decision. A variety of

information collected by different persons should result in better decisions and more effective interventions.

The sometimes difficult task in practical situations is to insure that different disciplines are involved not just as participants in the final staffing, but also as independent data collectors, observers, and interpreters of assessment information. The intent of a multidisciplinary team is to insure that a variety of disciplines and perspectives will be involved in the solution of a problem. Multidisciplinary teams which involve collection of data by only one or two of the team members fail to achieve the potential benefits of the multidisciplinary approach.

Outcomes Criterion. In the view of the present author the most useful definition of bias in assessment emphasizes the outcome of assessment activities for individuals. The fundamental and crucial question from this perspective is, What happens to a child as a result of assessment activities? If the child's competencies are improved and opportunities expanded as a result of the interventions that follow assessment activities, then assessment is beneficial, of high quality, and by definition, unbiased. Assessment that does not lead to interventions, or is followed by ineffective interventions is regarded as useless, and more likely to be biased.

The change proposed here, i.e., emphasis on outcomes, requires a departure from the training experiences and present orientation of some psychologists (see for example, Wade and Baker, 1977). The outcomes criterion deemphasizes the tradition of defining assessment as some kind of mysterious art in which the description of the "real" causes of behavior is the primary goal. The question of "What is really going on" which occurs frequently in case conferences usually leads to a description of the deep or underlying dynamics behind the behavior. In such discussions nothing is ever what it seems to be,

there is nearly always some symbolic significance of observed behaviors. Unfortunately these discussions are usually highly speculative with little or no objective verification of proposed hypotheses. These discussions usually lead to a description of the "pathology" presumed to be responsible for the behavior. Sometimes, juicy anecdotes about the reported sexual proclivities of family members are recited, followed by speculations on who views whom as what and how disturbed all that is and so on. These activities are unfortunately an important part of the tradition in applied areas of psychology. The alleged insight and understanding gained from such analyses are rarely translated into effective interventions, and in fact, probably deflect the attention of professionals from useful objective information. If the analysis of psychodynamic factors does not lead to effective interventions, and my impression is that it rarely does, then it should be regarded as simple voyeurism on the part of psychologists. Most importantly, it is poor assessment, and in many instances, biased assessment.

A further implication of the emphasis on outcomes of assessment is reduced level of inference. In order to achieve nonbiased assessment, i.e. assessment practices that lead to positive outcomes for individuals, we must gather information that translates directly into interventions. Part of the problem in our current practices is the use of fairly weak instruments in combination with "clinical insight" resulting in global descriptions of the person. An example from a psychological report might illustrate this point. A child made dark heavy lines in reproducing the Bender designs. The designs were reproduced accurately. However, the dark heavy lines were interpreted as revealing "repressed hostility." The clinical psychology literature and tradition is replete with similar interpretations (Rapaport, Gill and Schafer, 1968). The interpretations are <u>not</u> based on empirical data, but on analogical

reasoning. This reasoning assumes that some sort of logical relationship exists between observed behavior and underlying dynamics. Even when empirical data from groups support the interpretations, which is rare, the strength of the relationship is not sufficient to justify predictions for individuals. In fact, use of the clinical lore can usually be shown to result in more incorrect than correct decisions for individuals.

The problem is the level of inference. Many other examples could be cited, e.g. "Minimal" brain dysfunction, where global inferences are made on the basis of minimal data. There are several reactions to these clinical approaches, not the least of which is the increasing skepticism of the courts (Ziskin, 1975). Most important to our present concerns is the fact that assessment which results in a high level of inference is usually not related to interventions, and is therefore, of questionable benefit to individuals.

Placement Options and Effective Programs. If we accept the notion that possible bias in assessment is best conceptualized in terms of outcomes, then the availability of effective educational programs and alternative placement options is an absolute prerequisite to implementing nonbiased assessment procedures. In the situations which resulted in the special education placement litigation, the educational programs were apparently ineffective and the range of options limited. The author remembers all too well the very limited range of options that was typical until quite recently. The only choices often were regular classrooms with no assistance or self-contained, segregated classes for the mildly retarded. Psychologists can recall vividly cases where we knew the child was not "really" retarded, but in view of very low achievement accompanied by increasingly negative attitudes toward school and self, the self-contained, segregated class appeared to be the best option.

The situation has changed, or is in the process of change. A wide range of options is increasingly available, the principle of least restrictive alternative is the law of the land, and greater emphasis is placed on effectiveness of interventions through individualized educational plans with annual review. These are all positive changes. In the author's view, they are the best things that have happened for school psychologists in our history. However, school psychologists must participate with others in achieving the potential of these changes for improving interventions. Specifically, our assessment activities must be designed to yield information useful to choice of least restrictive alternative. Assessment must be directed toward the content of interventions, especially identifying specific areas of "educational" need in terms of social, emotional, and academic development. Assessment must also yield information concerning the approach to intervention, specifically changes in antecedent, situational, and consequent environments which can be used to carry out interventions. Finally, we need to gather information relevant to and/or assist others in evaluating the effectiveness of interventions. Discussion of some of the changes in assessment procedures that are underway or will become increasingly important in the future will appear in the next section.

Summary. Nonbiased assessment can only be achieved through insuring that effective outcomes result for individuals. In the view of the author, there are certain prerequisites to nonbiased assessment: (1) Procedural safeguards must exist. (2) Contributions of a multidisciplinary team including parents must be assured. (3) An outcomes criterion must be used. (4) Most importantly, placement options and effective programs must be available. Psychologists have important roles in achieving each of these prerequisites to nonbiased assessment. The combination of these prerequisites and the

specific changes in assessment practices to be discussed in the following section can produce assessment practices that are fair and useful for all children.

SPECIFIC CHANGES IN ASSESSMENT PRACTICES

Effective assessment in school psychology, i.e., assessment which leads to positive outcomes for individuals, is possible only within the broad context described in the previous section. These prerequisites to effective assessment must be met before the specific changes suggested in this section can be of benefit to individuals. If the educational system provides procedural safeguards, multidisciplinary resources, effective programs, and placement options, and if individual school psychologists are oriented toward an outcomes criterion, then more effective (and less biased) assessment can be achieved.

Problems in Classification of Children. Classification of children for the purpose of educational programming has been and will likely continue to be one of the important roles for school psychologists. This classification process inevitably involves some kind of labeling, and the debate over the alleged negative effects of labels has been very important in the litigation and legislation discussed previously. It appears that the controversy over labeling has subsided somewhat over the past few years for a variety of reasons. Hobbs (1975) recognized the necessity of some sort of classification (labeling?) in the final report of the project on classification of exceptional children. Hobbs described the dilemma facing special educators and school psychologists in the following terms:

> "Children who are categorized and labeled as different may be permanently stigmatized, rejected by adults and

> other children, and excluded from opportunities essential for their full and healthy development. Yet categorization is necessary to open doors to opportunity: To get help for a child, to write legislation, to appropriate funds, to design service programs, to evaluate outcomes, to conduct research, even to communicate about the problems of the exceptional child" (p. 3).

Efforts to "delabel" children or to completely avoid the possible negative effects of labels are probably destined to failure. There has been some suggestion that classification might be organized around the nature of the services needed, e.g., needs tutoring in reading, which would presumably have fewer negative connotations. The history of the meaning of other labels does not provide optimism about the effects of simply changing the basis for a classification system. Other labels which originally had precise, circumscribed meanings have over time been understood as global and perjorative.

Classification is inevitable and carries with it certain risks. In essence, when a child is classified by a multidisciplinary team an implicit, and now increasingly explicit, contract is established. Parents and children are asked to take the risk that the benefits from the services rendered as a result of the label will be substantially greater than the potential harm of the label (Gallagher, 1972). It is important that we recognize the risk that is involved with the label and take steps to minimize this risk. Even more important, effective services must be provided.

Classification: Minimizing the Risk. School psychologists are directly involved with intelligence testing and the classification of mild mental retardation. These are two areas in which the risk of misunderstanding is very high. One means to improve assessment is to devote efforts toward clarifying the meaning of current procedures which are known to carry high risk. A clarification statement regarding the meaning of IQ test results appeared in an earlier section. Clarification of the meaning of the term mild mental

retardation is probably more difficult, but no less important. The original meaning of mild mental retardation intended by the American Association on Mental Deficiency (AAMD) (Heber, 1961) is quite different from the way it is commonly understood today. In fact, mild mental retardation referred originally to the current status of the individual with no presumptions about etiology or future status. In Mercer's (1978) terminology, it was a social system classification. Whether the term was ever understood correctly by most is questionable; it certainly is misunderstood by the public and many professionals today. Perhaps, mild mental retardation is a somewhat unique problem within the special education classification system due to the nature of the AAMD classification and terminology system. The 1961 AAMD system, revised slightly in 1973 and 1977, was quite properly organized around the dimensions of adaptive behavior and intelligence. Unfortunately, the AAMD system included the implicit assumption that quantitative factors were the major differences between levels of mental retardation. Thus, the same general classification, mental retardation, was applied to persons who had vastly different etiologies, competencies, and prognoses. The factors of etiology, current behavior, and prognosis are partially accounted for in the AAMD system by the adjectives of mild (educable), moderate, severe, and profound. Unfortunately, the distinctions suggested by these adjectives have usually not been understood by persons not directly involved with exceptional children.

Two steps should be taken to minimize the risks associated with classification. First, the classification terminology used should be descriptive, behavioral, and reflect a low level of inference (see previous discussion). The general change toward use of the term learning disability rather than neurological dysfunction or minimal brain injury is consistent with these principles. Use of the term behavior disorder rather than emotional distur-

bance is another example. A similar change is needed in the mental retardation terminology and classification system, and perhaps, in other areas as well. Most of the mildly retarded, regardless of race or ethnicity, are so called "Six Hour Retarded" children. Their difficulties are fairly specific to the public school situation. A change in the classification terminology which would reflect more accurately the nature of mild mental retardation is needed. A term such as general academic disability might be less perjorative, less likely to be misunderstood, and less likely to be associated with more severe forms of mental retardation. It is important to note that changing the terminology will not avoid all of the stigmatizing effects of labels. Even when completely neutral terms were adopted, e.g., recall the term "garden variety", the terms become associated with negative connotations fairly quickly. In the case of mental retardation, especially the educational use of the term, a change in terminology would reduce, but not eliminate, the risks associated with classification.

A second, more practical, approach to reducing the risks associated with classification is to clarify the meaning of the terminology. Mild or educable mental retardation does not mean comprehensive and permanent incompetence of biological origin. In fact, most mildly retarded persons achieve relatively normal lifestyles as adults. Do we communicate these facts to parents, students, and other professionals in an effective manner? When working as a school psychologist I had the good fortune of being associated with a very capable director of special education. We were concerned about misconceptions regarding the term educable mentally retarded. Jointly we developed a question and answer "fact sheet" on educable mental retardation which was then provided to every teacher (regular and special), parents, principals, and in the case of older students, to every student in the special class program. Similar

attempts to clarify the meaning of other special education categories are probably needed. These efforts will not eliminate all of the risk associated with classification, but they may reduce these risks.

Classification: Maximizing the Benefits. One of the major underlying assumptions in the special education litigation was the presumed ineffectiveness of special education programs for the mildly retarded. Assessment is biased according to the definition used in this paper if effective interventions are not provided. Perhaps the most important step in reducing bias and maximizing the benefits of classification is to insure effectiveness in the interventions that follow. Cromwell (1975) provided a simple, but important model for analyzing the usefulness of diagnostic constructs. The crucial feature of this model is the requirement that classifications based on etiological information or current behaviors (called A and B by Cromwell) must be related to information on outcome (D) or information on interventions and outcomes (C and D) in order to be useful. Diagnostic constructs based only on AC or BC types of information were regarded as useless since they only relate diagnosis to currently available or popular treatments which are of unknown benefit to persons.

The kind of information collected by school psychologists during initial assessment, and the contributions of all persons on the staffing teams should be oriented toward designing interventions and evaluating their outcomes. Obviously, classification, even if conducted with great skill, is not sufficient. Intellectual assessment, while important to classification, provides limited information about the kind and nature of interventions. School psychologists along with others on the staffing team must collect information that is directly relevant to specific educational need. All of the members of the staffing team share the responsibility of using a broad base of assess-

ment information to select program options, intervention goals, intervention strategies. School psychologists have much to contribute to each of these decisions. A current concern is the degree to which school psychologists are oriented toward and willing to participate in planning, evaluating, and in certain instances, conducting interventions. Limited anecdotal evidence suggests that at least some school psychologists see these activities as beyond the scope of their role. The current requirements of multidisciplinary staffing, individualized educational plan (IEP), annual review of educational program, and review of classification every three years provide excellent opportunities for school psychologists to work with others in insuring the effectiveness of programs. Furthermore, school psychologists should play a special role in actually carrying out or working very closely with those who implement interventions designed to improve social or emotional competencies. Interventions in the areas of social or emotional adjustment should also be designed by the multidisciplinary team, IEPs developed, interventions evaluated, and so on.

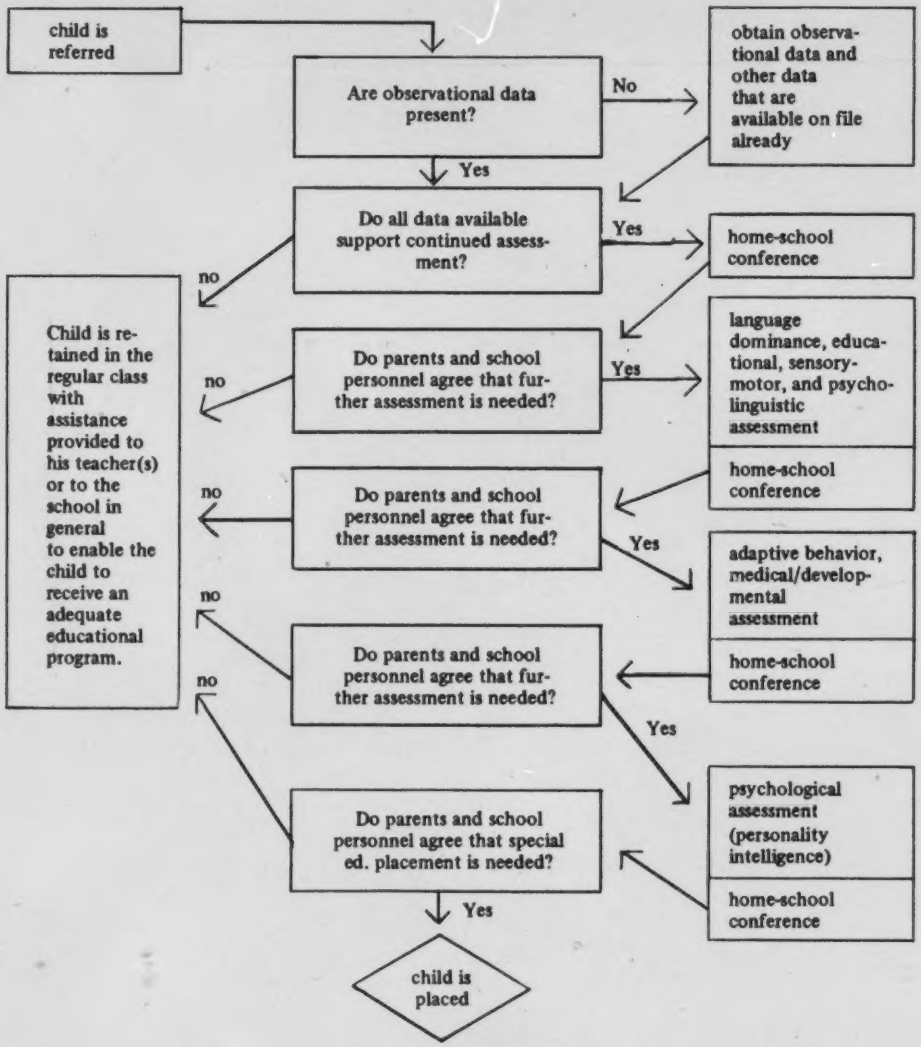
Summary. Overall, school psychologists along with other personnel have crucial responsibilities in insuring the effectiveness of interventions. If interventions are effective in improving competencies and expanding opportunities, and if the risks associated with classification are reduced, the special education bargain is a good one for parents and children regardless of racial or ethnic factors. If these standards are not met, classification and the special education contract is a bad bargain and likely to be biased. Recognition of and coping with the risks involved with classification is an important step toward improving the probability of fair and effective assessment.

Multifactored Assessment. The concept of multifactored assessment was the apparent solution to the dilemma of defining and describing the requirement of nonbiased assessment. The requirement of multifactored assessment is suggested in the 94-142 rules and regulations, and is even more prominent in the reports from the CORRC project. The underlying assumption is that assessment is likely to be less biased if a broad variety of information is collected and considered systematically in making placement decisions. In the view of the present author this assumption is sound, but insufficient. Improved classification decisions are certainly important, but even more important is the use of the multifactored information in designing and evaluating interventions.

Tucker (1975, also in Oakland, 1977) described the categories of information which should be developed in a comprehensive assessment of children "for possible mildly handicapping conditions." For the most part, the categories of information are fairly standard and consistent with traditional descriptions of comprehensive psychoeducational evaluations. The arrangement of the categories of information, especially the sequence suggested for collecting the information is somewhat unique (See Figure 1 reprinted from CORRC report). Especially noteworthy is the placement of "psychological assessment" (personality and intelligence) at the end of the sequence of assessment procedures.

The nine categories of information needed for a multifactored assessment according to Tucker (1975) were: (1) Observational data which are gathered for the purpose of determining the degree of deviance (if any) of the child's behavior in relation to other children in the same environment; (2) Other data available, e.g. records of previous performance, which are used to corroborate or contradict the deviance established in step 1; (3) Language dominance data

Figure 1
COMPREHENSIVE INDIVIDUAL ASSESSMENT
For Possible Mildly Handicapping Conditions


child is referred
Are observational data present?
No
obtain observational data and other data that are available on file already
Yes
Do all data available support continued assessment?
Yes
home-school conference
no
Child is retained in the regular class with assistance provided to his teacher(s) or to the school in general to enable the child to receive an adequate educational program.
Do parents and school personnel agree that further assessment is needed?
Yes
language dominance, educational, sensory-motor, and psycholinguistic assessment
home-school conference
no
Do parents and school personnel agree that further assessment is needed?
Yes
adaptive behavior, medical/developmental assessment
home-school conference
no
Do parents and school personnel agree that further assessment is needed?
Yes
psychological assessment (personality intelligence)
home-school conference
no
Do parents and school personnel agree that special ed. placement is needed?
Yes
child is placed

Reprinted from Tucker, J. Operationalizing the diagnostic-intervention process. in Nonbiased assessment of minority group children: With bias toward none. Lexington, Ky.: Coordinating Office for Regional Resource Centers, 1976. (p. 46).

which are used to determine the appropriate language for further assessment and as information which may influence the interpretation of data collected in previous phases; (4) Educational assessment data (usually collected with individually administered devices) which are used as further evidence of extent and nature of problem and for programming; (5) Sensory-motor and/or psycholinguistic assessment data which according to Tucker should be used in consideration of placement in learning disabilities programs and possibly, in educational programming generally; (6) Adaptive behavior data which are apparently used for classification, and perhaps for programming; (7) Medical and/or developmental data which are used to rule out a medical etiology of the problem as well as to provide a basis for referral to appropriate medical services if needed; (8) Personality assessment data including self-report which are used to determine the degree (if any) of emotional involvement; and (9) Intellectual assessment data which are used "to estimate the level of a child's intellectual functioning."

Implicit in the above description of multifactored assessment is the concern about misuses of IQ test information. The placement of intellectual assessment data at the <u>end</u> of the sequence appears to be a conscious effort to emphasize the importance of other data. The other data were seen as important both for the interpretation of IQ test results and for decisions about special education placement. Interestingly, the area of sociocultural background was not mentioned as a separate category in Tucker's description of comprehensive (multifactored) assessment. However, concerns about sociocultural background are implicit in the descriptions of language dominance and adaptive behavior. Sociocultural background is mentioned as a separate area of assessment in the 94-142 rules and regulations, and should be recognized as an important area in the multifactored assessment.

Summary. Multifactored assessment is crucial to appropriate classification and to effective interventions. The approach suggested by Tucker is especially useful in that the relationship of other data to IQ test results is clarified. Most of the categories of information suggested by Tucker are fairly standard and will therefore, not be discussed further in this paper. Two categories of information, primary language and adaptive behavior, are less commonly included in comprehensive assessment procedures. These categories of information along with sociocultural background data will be discussed in sections that follow.

Multifactored Assessment: Primary Language. The assessment of primary language competence is a logical, common sense procedure as well as a requirement in the recent legislation. Non-English speaking children have apparently been placed in programs for the mildly retarded on the basis of tests administered in English (see Diana or Guadalupe cases). These classification and programming decisions were inappropriate, although an even larger problem in those situations was the apparent absence of alternative programs for Non-English speaking youth.

Assessment of primary language competence is more difficult than it might appear. Many instruments have been developed recently (see Oakland, 1977), but little systematic work has been conducted on their reliability and validity. Nevertheless, systematic effort to assess primary language competence is needed. The decision about primary language competence must be based on data. The presence of a Latino surname, for example, is certainly not sufficient to conclude that the child or family uses Spanish as the dominant language. The author is acquainted with cases of Latino surnamed families where Spanish is not spoken, and has not been used in the family for several generations. Conversely, the author encountered a case in 1967 in

eastern Iowa where the child had an Anglo surname, but was monolingual Spanish speaking.

The information on primary language is important in collecting and interpreting other assessment data, and in decisions about appropriate interventions. If the child is monolingual, Non-English speaking, perhaps the wisest course of action is to simply avoid the use of norm referenced standardized tests of achievement and ability. The 94-142 regulations suggest use of an interpreter. Due to the many problems which arise when attempts are made to translate tests into other languages, e.g. items do not have the same meaning and difficulties of items change, the results of translated tests are of questionable value. If inferences must be made about ability, use of nonverbal or performance tests is probably the best course of action. Educational programs for monolingual Non-English speaking students must be provided in the students' native language if at all feasible (Lau vs Nichols). If only a few monolingual children attend schools in a particular district, then other alternatives should be pursued (see Oakland, 1977).

Bilingual children may exhibit widely varying competencies in English and another language. The range will extend from limited to high degrees of competence in either or both languages. The language dominance measure that is used to determine primary language should be supplemented by other measures which yield information on competence in both languages. Subsequent assessment activities should be conducted within the dominant language of the child. An important principle to remember is the assumption of maximum performance. Any inference about ability or academic aptitude made in subsequent assessment activities should include consideration of the effects of differences in language. Bilingual youth may, though certainly not always, obtain lower scores on verbal measures administered in English due to limited exposure to

English. Special education services may not be the appropriate intervention for bilingual children who, on the basis of other data, meet the state guidelines for special education classification. Bilingual/bicultural programs may be more appropriate, and children's rights to such services have been established through the Lau decision.

Multifactored Assessment: Adaptive Behavior. A subtle but important change has occurred over the past fifteen years concerning the criteria for judging adaptive behavior competence among school age children. The 1961 AAMD definition of mental retardation suggested that learning and school achievement were the principle criteria for assessing adaptive behavior for school age children. These criteria have been expanded through litigation, legislation, federal memoranda, (e.g. Office of Civil Rights, 1972), and now recently, by the AAMD in the classification and terminology manual (Grossman, 1973, 1977). Adaptive behavior for school age children now encompasses activities outside of the school including the settings of home, neighborhood, and community. Clearly, the criteria have changed, and it is no longer appropriate to judge adaptive behavior on the basis of school performance alone.

Conceptions of and methods for measuring adaptive behavior have been fairly limited until quite recently. Concepts of adaptive behavior have usually been restricted to fairly simple self-help or social behaviors such as dressing, eating, etc. These behaviors are usually mastered by normal and mildly retarded children prior to or soon after school entrance. Most of the instruments available currently reflect these limited conceptions, and are not particularly useful for most normal or mildly retarded children. Perhaps the best example is the AAMD Adaptive Behavior Scales which were developed from careful studies of deficit behaviors among samples of persons in institutions for the mentally retarded, and later normed on samples

selected from institutions for the mentally retarded. A public school version of the AAMD Adaptive Behavior Scales was developed recently (Lambert, Windmiller, and Cole, 1975). The usefulness of the Public School Version is questionable since all of the items were selected from the original AAMD version with norms developed on school age children. The respondent in the AAMD Public School version is the classroom teacher, and the content validity of the items is questionable. Similar criticisms apply to the other adaptive behavior scales, nearly all of which were developed through studies of the more severely retarded.

A major advance provided by the System of Multicultural Pluralistic Assessment (SOMPA) (Mercer and Lewis, 1978) is the development of the Adaptive Behavior Inventory for Children (ABIC) which is designed for normal children and is appropriate for the mildly retarded. The emphasis is on adaptive behaviors outside of the school setting. Adaptive behavior is conceptualized as the degree to which the child performs increasingly complex social roles. The 242 items were organized on a judgmental basis into the social role categories of family, community, peer relations, nonacademic school, earner/consumer, and self-maintenance. Standard scores and age graded norms are provided for each category and for total score. Responses to the items are provided by the primary "caretaker" of the child, usually the mother. The norms for the ABIC are based upon data collected from a representative sample of California school age children between the ages of five and eleven. The accuracy of the ABIC norms for children in other geographic regions has not yet been determined, and will undoubtedly be the focus of considerable research in the next few years. I suspect the norms will hold up reasonably well, but some caution should be exercised in interpreting ABIC data until definitive studies are available.

Even though much research with the ABIC needs to be conducted, cautious use of the instrument in classification and programming decisions seems justified by the preliminary data presented by Mercer and Lewis (1978) and Mercer (1973). Perhaps the most important use of the ABIC data is for selection of program option rather than determination of original classification. The original intent of the senior author of the ABIC was apparently to broaden conceptions of and then provide a measure of the dimension of adaptive behavior. In other words, the primary purpose was to improve accuracy in the classification of mild mental retardation. Children who were referred for psychological evaluation due to difficulties in school performance were to be classified as mildly retarded only if they failed both the adaptive behavior and the intellectual dimensions of the AAMD definition of mental retardation. Children who met the criterion of low IQ, but obtained scores in the normal range on the adaptive behavior measure were termed "Quasi-retarded" in an earlier conceptual scheme proposed by Mercer (1973). Classification of "Quasi-retarded" children as mildly retarded and placement in school programs for the educable mentally retarded was seen as inappropriate.

Mercer (1978) has apparently revised her position on whether or not the "Quasi-retarded" should be served in special education programs. The key issues are the type of program used and the effectiveness of the program. Programs which carry relatively higher risks (see previous discussions) such as special classes are seen as less appropriate than programs with possibly lower risks that have equal or perhaps greater benefits, e.g. resource programs.

The degree to which the ABIC data are used in classification decisions will probably vary depending on state education codes and local policies. Children who obtain IQ scores in the mildly retarded range, but are normal in terms of adaptive behavior in the home and neighborhood are very often in need

of some kind of special services. The important question for us is not simply whether to classify the child, but what kind of program is needed.

Information from the ABIC or similar data from other sources should be used to select service options. The curriculum in special classes for the educable mentally retarded has traditionally reflected emphases on social competence and functional academic skills. The emphasis on social competence is justifiable in view of the longitudinal data on the adult adjustment of persons who are mildly retarded. The problems of mildly retarded persons in vocational settings are more likely to be related to deficits in social competence than limited academic or intellectual skills. However, the ABIC data or data from other sources may confirm that the child has relatively high social competence in the home, neighborhood and community. If social competence is relatively normal, the traditional special class is probably not the best service option. An educational program which is more specific to the child's academic needs such as the resource option is probably more appropriate. The "Quasi-Retarded," i.e. children with low IQs and normal adaptive behavior, should be served in resource programs in most, perhaps nearly all cases.

The use of the adaptive behavior data proposed here, i.e., selection of service option, is consistent with current special education rules and regulations for most states. In these guidelines classification or eligibility is described independent of service option. Unfortunately, there seems to be an erroneous assumption that service option is determined, or at least heavily influenced, by special education category. It seems that many persons view special classes as the only option for children with mild mental disabilities. In fact, the full range of service options including the resource approach should be available to children with mild mental disabilities.

In addition to selection of service option, another important use for adaptive behavior data is educational programming. Deficits in social competence should be viewed as potential goals for educational programming. The kind of data that are provided by Mercer's ABIC is probably not sufficient for precise specification of goals for interventions in social competence. The ABIC or other data should be useful for determining whether there are problems with social competence and the general nature of these problems, if any. These data would need to be supplemented by more precise observation prior to designing an intervention.

A number of conceptual and practical issues concerning the assessment of adaptive behavior are discussed by Coulter and Morrow (1978). The discrepancy between what is needed or required in this area and present technology is rather large. Considerable progress in instrument development is needed. In addition to the problems related to currently available instruments, a number of conceptual issues must be resolved. Is adaptive behavior data useful for classification decisions in educational settings? Is adaptive behavior data useful in program planning for the mildly handicapped? What is the nature of adaptive behavior in older age groups? The reader is encouraged to consult Coulter and Morrow's discussion of these and other issues.

Summary. Adaptive behavior information will become an increasingly important component of the multifactored assessment of children suspected of having mild handicaps. The degree and nature of the use of adaptive behavior data in educational classification decisions with the mildly retarded in the future are not entirely clear. However, these data appear to have high potential for use in decisions concerning selection of service option. Furthermore, adaptive behavior data from instruments such as the ABIC may provide important

preliminary information for decisions about interventions in the area of social competence.

Multifactored Assessment: Sociocultural Background. Recent federal regulations require the consideration of sociocultural background in special education placement decisions. Unfortunately the federal regulations do not suggest procedures either for measuring or using the sociocultural background data. The CORRC Project on nonbiased assessment did not deal directly with the issue of measurement and use of sociocultural data. This section of the paper will consider the concept of sociocultural background, its relationship to intelligence and achievement, and procedures for measuring and using sociocultural background data.

The concept of sociocultural background includes the overlapping factors of social class and race or ethnicity. Mercer (1978) refers to the concept of "eth-class" which is a term from sociology that refers to the combined effects of ethnicity and socioeconomic status. The concept of eth-class, or the more commonly used term sociocultural background, is needed to accurately describe the relationship between sociocultural factors and achievement or intelligence. In fact, social status and racial or ethnic background are NOT independent in the population of the United States. Specific racial or ethnic groups are consistently under or overrepresented in high or low social statuses. Specifically, the incidence of poverty is much higher among groups such as Blacks, Latinos, Native Americans, and Appalachian Whites. Thus, these groups differ on the average from the rest of the population on measures of social status.

The concept of socioeconomic status (SES) would be sufficient for our discussions if all racial or ethnic groups of the same social status performed in the same way on measures of achievement and intelligence. This, however,

is not the case. In addition to and independent of social status, racial or ethnic factors influence performance on achievement and intelligence tests (e.g., Lesser, Fifer, and Clark, 1965; Reschly, in press). The results of these and other studies indicate that although social status influences the level of performance for all groups, ethnicity or race influences the pattern of performance. Thus, social status or ethnicity or race alone are insufficient to account for their combined effects on achievement or intelligence.

The relationship of sociocultural background to measures of achievement or intelligence is far from perfect. In fact, some low SES minority children obtain very high scores on measures of achievement and intelligence, and conversely, some high SES Anglo children obtain low scores. The relationship of the sociocultural factors to average levels of performance appears to be more impressive. For example, Kaufman and Doppelt (1976) reported differences of nine to seventeen points both for blacks and whites between the highest and lowest SES groups in the WISC-R standardization sample.

The relationship of sociocultural factors or eth-class to achievement and intelligence has been explained in a variety of ways. The controversies surrounding these explanations have been one of the crucial, but usually unrecognized, factors that led to the litigation over special education placement. We might speculate, for example, about the effects on this litigation of Jensen's hypotheses of hereditary differences among racial groups. It should be emphasized that this is speculation, but the movement to ban IQ tests might never have occurred if it were not for the widespread publicity accorded Jensen's views. It would be highly ironic if IQ tests are banned through the Larry P. decision because of the views of one of the strongest proponents (Jensen) of IQ tests. This outcome is unlikely, but the sequence of events leading to the case is important to understand. Even more important

for users and consumers of IQ test results are the facts that IQ tests do indeed measure learning (not innate potential), and IQ test results are influenced by sociocultural factors as well as heredity. Other explanations for the IQ test differences between sociocultural groups exist, and are supported by data (e.g. Garber, 1975). The explanations for IQ test differences have influenced the special education litigation, and will influence procedures to develop nonbiased assessment practices.

A variety of measures of social status have been used in the published research. The measures vary from rather simple indices such as Duncan's occupational scale to more complex measures such as the Warner Index. The Duncan occupational scale has been the most frequently used measure since the information needed is usually obtained easily. The more complex measures have been used less frequently, but are correlated at a slightly higher level with achievement or intelligence. The adequacy of the social status measures for equating groups has been questioned recently (Trotman, 1977). Matching groups through statistical analyses or selection of subjects on social status, (Jensen's approach), does not equate groups on all, or apparently, even the most important background factors related to achievement or intelligence. Trotman reported data on the relationship of a home intellectuality scale to achievement and intelligence. Groups of black and white students equated on social status differed substantially on the home intellectuality scale. (Home intellectuality was measured by a 63 item questionnaire and rating scale designed around environmental variables related to intelligence).

The Sociocultural Measures (SCM) of the SOMPA provide a relatively complex measure of sociocultural background. The SCM scales were built around variables that have been correlated with IQ results in published studies. Factor analysis results were used to organize the 24 SCM items into nine

factors and then into four sociocultural measures. The four sociocultural measures that resulted were named Urban Acculturation, Socioeconomic Status, Family Structure, and Family Size. Two characteristics of the SCM should be noted. First, the measures are considerably more complex in comparison to the traditional indices of social status. Secondly, the measures are focused on relatively unchangeable characteristics of the child's family background. In contrast to the data from, for example, a home intellectuality measure, there is relatively little that could be done in terms of intervening or changing the conditions measured by the SCM.

In the SOMPA the SCM are used in a number of ways. First the SCM data are used to describe the child's position in "sociocultural space," essentially the combination of social status, degree of participation in the dominant culture, and degree of similarity of family to the dominant culture in terms of attitudes, values, and lifestyle. Second, the SCM results for a specific child are compared to the average SCM results for Anglo children. This comparison is seen as an index of social distance between the child and the school which is believed to represent Anglocentric, middle class values. Third, and most important, the SCM are used to predict the expected IQ score for children with specific sociocultural characteristics. If the predicted score is <u>below</u> the population mean, then the individual's score is adjusted upward based upon a comparison of the obtained score to the predicted score. The adjusted score is called the Estimated Learning Potential (ELP).

The meaning and usefulness of the ELP score will undoubtedly be the source of considerable controversy in the next few years. The ELP score is based upon the relationship of the SCM to IQ scores for groups (Anglo, Black, and Hispanic) and then applied to individuals. The relationship of the SCM to WISC-R scores was determined through multiple regression analyses. The

SCM accounted for 28%, 14%, and 24% of the variance in the Verbal, Performance, and Full Scale WISC-R IQ scores when the data for all three groups were combined. When the data for the three groups were analyzed separately, the amount of variance accounted for was diminished (range of 10% to 22% on Verbal Scale, 4% to 11% on Performance Scale and 14% to 18% on Full Scale). The multiple correlations, reflecting the relationship of the four SCM with the WISC-R varied from .19 to .47 (median was .37).

The group specific regression equations are then used to predict an "expected" WISC-R IQ score for the individual. If the individual's expected score is 100 or above, the ELP is simply the obtained WISC-R score. If the predicted score is less than 100, indicating that the child's sociocultural background is discrepant from the Anglocentric, middle class school, the score is adjusted. The adjusted score is based on a comparison of the obtained score to the predicted score and then transformed to a score scale with a mean of 100 and standard deviation of 15, i.e., to a conventional IQ score scale. This adjusted score is interpreted as estimated learning potential.

Prior to discussion of the meaning and use of the ELP score, certain limitations must be mentioned. First, the ELP scores in the SOMPA are based entirely upon data collected in California. The accuracy of these scores in other settings will depend upon the similarity of the regression equations across different settings. Some preliminary data suggest that although the weights of SCM vary across settings, the obtained ELPs are similar, but not identical (Oakland, 1977). Features of the regression equations other than the weights may also vary. If the constant (also called the intercept) or the multiple correlation vary substantially from the California results, the ELPs will also vary. Again, we must wait for additional studies before reaching conclusions concerning the applicability of SOMPA norms to other settings.

More fundamental concerns about the ELP exist regarding reliability and validity. An unreliable measure cannot be valid. (Anastasi, 1976). The ELP scores for children with predicted scores of less than 100 are based on a multiple regression equation. As mentioned previously, the obtained score is compared to the predicted score and then transformed to another score scale. Linear transformation of a set of scores does not of course affect the reliability of the scores. The reliability of scores predicted from another set of scores is directly proportional to the multiple correlation. In the case of the SOMPA ELP scores, the reliability of the ELP scores is directly proportional to the multiple correlation of the SCM and WISC-R scores. The size of the multiple correlations and the associated standard errors of estimate do not provide confidence in the reliability of the ELP. The ELP in the view of the present author is a rather unreliable score (Note: Additional information on this question will be provided in the first issue of the 1979 volume of The School Psychology Digest).

If the ELP is a fairly unreliable score, the potential relationship to other measures is of course severely limited. Mercer has suggested that the appropriate criterion for examining the validity of the ELP is the rate of acquisition of new information or skills. Budoff's (1975) procedure for measuring learning potential is one of the possible methods for gathering data on acquisition rate. One problem with Budoff's procedure is the extensive amount of time required to determine acquisition rate. Due to the time apparently involved with studies of acquisition along with the problem of selecting appropriate subjects and materials, few studies of the kind suggested by Mercer are likely to be undertaken. However, such studies may not provide strong support for the validity of the ELP due to the problems with reliability.

A number of studies investigating the relationship of the ELP to conventional measures of achievement will undoubtedly appear in the next few years despite Mercer's objections. Mercer has suggested that conventional measures of achievement are not appropriate criteria for examining the validity of the ELP. One such study involving three sociocultural groups, (Oakland, 1977), reported higher correlations for the conventional IQ score (median = .64) in comparison to the ELP (median = .48). The criteria for achievement were the reading and mathematics scores from the California Achievement Test. Other studies reporting similar findings will probably appear in the near future. These results, although not directly relevant to the construct of ELP, will provide some useful information regarding the meaning of the ELP.

Perhaps it is important to separate the construct of ELP from considerations of its predictive validity. The construct is important. The SOMPA and the ELP represent, among other things, an attempt to emphasize the importance of sociocultural factors in the intellectual or academic performance of children. The different perspectives suggested in SOMPA are extremely important. The specific suggestion of regarding the WISC-R conventional scores as social system measures is certainly appropriate in terms of the massive body of knowledge on the relationship of various factors to intelligence. Renaming of the conventional IQ scores as School Functioning Level (SFL) is consistent with the validity evidence for IQ tests. IQ tests do measure, although imperfectly, the likelihoodof success in educational settings. This information is important, but limited. The SOMPA approach will make an important contribution toward recognizing the limitations of IQ test results.

Summary. Consideration of sociocultural background is required by recent federal regulations. Mercer's SOMPA provides one of the more complex methods of measuring sociocultural factors and the only procedure known to the author

for using these data systematically in interpreting IQ test results. However the use of the sociocultural data to adjust WISC-R scores in the SOMPA will be criticized on a number of bases. The predictive validity of the ELP is doubtful. However, the construct of ELP and the use of sociocultural data in understanding the meaning of IQ test results are valuable contributions of the SOMPA.

NONBIASED ASSESSMENT: SOME TENTATIVE CONCLUSIONS

Nonbiased assessment is obviously an extremely complex issue. Concerns with the meaning and usefulness of IQ test results have dominated much of the discussion of nonbiased assessment. The issues surrounding the meaning of IQ (academic aptitude) have been debated for at least sixty years, and are not likely to be resolved in the near future. However, many other issues such as the meaning and etiology of mild mental retardation, the rights of parents and students, the effectiveness of special education interventions, and the definition of bias in tests are clearly involved with our efforts to reduce bias in assessment. These issues have been discussed in this paper, though certainly not resolved.

There are two possible reactions among a range of possible reactions by school psychologists to the pressures for nonbiased assessment which could be damaging to children. One possible reaction is to conclude that the issue is so complex and ill defined that there is nothing we can do, hence, we should stubbornly defend and simply continue our current practices. This reaction will be maladaptive. There are important changes that we can make which will enhance the fairness and usefulness of assessment for all children. In the interests of children, we need to make these changes. A second

maladaptive reaction is to reject most if not all of our current instruments and practices. For example, some have rejected the use of IQ tests with culturally different children. Others have severely limited the numbers of culturally different children in special education programs simply on the basis of their proportions in the population. Such reactions are not in the best interests of children.

Positive reactions to the concerns about nonbiased assessment must first be based on a recognition of the ambiguity of the current situation. There are no and probably never will be any easy solutions.

Recognition of the underlying assumptions in the special education placement litigation provides an orientation to the most important issues in nonbiased assessment. One can only wonder if these cases would have appeared IF the interventions were effective; IF due process safeguards had been observed; IF the interventions had been consistent with the principle of least restrictive alternative, i.e., had not been provided in segregated, self-contained special classes; IF the assessment had been multifactored and programs based on specific educational need; and so on. The fact is that assessment by school psychologists and programs in special education did NOT meet these criteria in at least some, and perhaps, many instances. The litigation and legislation are attempts to correct these abuses. From the perspective of school psychologists, the current demands for nonbiased assessment along with the other requirements from the courts and legislation, are the best things that have happened for our profession (and for children).

Three general themes should form the basis for efforts to achieve nonbiased assessment. First, and most important, we must continue and expand our efforts to insure that assessment procedures result in positive benefits for individuals. This goal is certainly not new. The underlying assumption

of positive benefit to individuals has always been the goal in all types of assessment. Realization of this goal requires more concern about the relationship of our assessment activities to interventions, and more concern about the effectiveness of these interventions.

A second theme is the need to implement the idea of multifactored assessment. Again, this is not a new idea. However, the degree to which comprehensive assessment was conducted, documented, and used in planning interventions has varied considerably. The proper role of IQ tests in the multifactored assessment must be recognized. Areas often ignored in the past, e.g., adaptive behavior outside of school, primary language competence and sociocultural background, should be a part of the assessment process. These newer areas of assessment, along with the conventional areas, are important to better understanding of children. Fuller understanding can lead to better, more refined classification decisions and more effective interventions.

Finally, our understanding of nonbiased assessment and our ability to implement these procedures will be enhanced if we view nonbiased assessment as a process rather than a set of instruments. The process is oriented toward insuring fairness and effectiveness of assessment and interventions for all children. The process is appropriate in all settings regardless of the ethnic or racial composition of the student population. The nonbiased assessment process is perhaps best illustrated by the series of questions developed by the Northeast Regional Resource Center. A copy of this document is included in an appendix. A second guideline which follows was developed by a committee appointed by the Iowa Department of Public Instruction. Both documents are attempts to identify key features of a nonbiased (and effective) assessment process.

REFERENCES

American Psychological Association Monitor. Exhibit A: IQ Trial. Plaintiffs Take the Stand. p. 4, December, 1977.

Anastasi, A. Psychological testing. (4th ed.) New York: Macmillan, 1976.

Ashurst, D. & Meyers, E. Social system and clinical model in school identification of the educable retarded. In Eyman, R., Meyers, E., and Tarjan, G. (Eds) Sociobehavioral studies in mental retardation. Washington, D.C.: American Association on Mental Deficiency, 1973.

Abeson, A., Bolick, N. & Hass, J. A primer on due process. Reston, VA: Council for Exceptional Children, 1975.

Bakke vs. Regents of the University of California.

Baller, W., Charles, D. & Miller, E. Mid-life attainment of the mentally retarded. Genetic Psychology Monographs, 1967, 75, 235-329.

Bernal, E. A response to "Educational uses of tests with disadvantaged subjects". American Psychologist, 1975, 30, 93-95.

Block, N. & Dworkin, G. The IQ controversy. New York: Pantheon/Random House, 1976.

Brody, E. & Brody, N. Intelligence: Nature, Determinants, and Consequences. New York: Academic Press, 1976.

Budoff, M. Measuring learning potential: An alternative to the traditional intelligence test. In Gredler, G. (ed.) Ethical and legal factors in the practice of school psychology. Harrisburg, PA: Pennsylvania Department of Education, 1975. (Also, see Budoff et. al. article in the American Journal of Mental Deficiency, 1971, 76, 159-169.

CORRC Coordinating Office on Regional Resource Centers. With Bias Toward None. Lexington, KY: University of Kentucky, 1975.

California State Board of Education News Release, January 15, 1975.

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Cleary, T. A., Humphreys, L., Kendrick, A., & Wesman, A. Educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-41.

Coulter, A. & Morrow, H. The concept and measurement of adaptive behavior. New York: Grune & Stratton, 1978.

Cromwell, R., Blashfield, R. & Strauss, J. Criteria for classification systems. In N. Hobbs (ed.) Issues in the Classification of Children. San Francisco: Jossey-Bass, 1975.

Darlington, R. Another look at "cultural fairness". Journal of Educational Measurement, 1971, 8, 71-82.

Davis, A. & Eells, Davis-Eells Tests of General Intelligence or Problem Solving Ability. New York: Harcourt, 1951.

Dent, H. Assessing black children for mainstream placement. In Jones, R. (ed.) Mainstreaming and the minority child. Reston, VA: Council for Exceptional Children, 1976.

Diana vs. State Board of Education, C-70 37 RFP, District Court for Northern California (February, 1970).

Dove Counterbalanced Intelligence Test. Mimeo, undated.

Dunn, L. Special education for the mildly retarded: Is much of it justifiable? Exceptional Children, 1968, 35, 5-22.

Edgerton, R. The cloak of competence: Stigma in the lives of the mentally retarded. Berkeley, CA: University of California Press, 1967.

Education for all Handicapped Children Act of 1975. Federal Law 94-142.

Foster, G. & Ysseldyke, J. Expectancy and halo effects as a result of artificially induced teacher bias. Contemporary Educational Psychology, 1976, 1, 37-45.

Gallagher, J. The special education contract for mildly handicapped children. Exceptional Children, 1972, 38, 527-535.

Galton, F. Hereditary Genius. London: Macmillan, 1892.

Garber, H. Intervention in infancy: A developmental approach. In Begab, M. & Richardson, S. (Eds.) The mentally retarded and society: A social science perspective. Baltimore: University Park Press, 1975.

Gerken, K. Assessment of high risk preschoolers and children and adolescents with low incident handicapping conditions. In Phye, G. & Reschly, D. (Eds.) School psychology: Perspectives and Issues, New York: Academic Press, in press.

Grossman, H. (Ed.) Manual on terminology and classification in mental retardation. Washington, D.C: American Association on Mental Deficiency, 1973. Also, 1977 revision.

Guadalupe vs. Tempe Elementary School District, 71-435, District Court for Arizona, January, 1972.

Haywood, H. C. Some perspective on social-cultural aspects of mental retardation. In Haywood, H. C. (Ed.) Social-cultural aspects of mental retardation. New York: Appleton, 1970.

Heber, R. A manual on terminology and classification in mental retardation (2nd ed.) American Journal of Mental Deficiency, 1961, Monograph Supplement 64.

Heber, R. et. al. Rehabilitation of families at risk for mental retardation. Madison, WI: Rehabilitation Research and Training Center in Mental Retardation, University of Wisconsin, 1972.

Hewett, F. & Forness, S. Education of exceptional learners. Boston: Allyn and Bacon, 1974.

Hobbs, N. The futures of children. San Francisco: Jossey-Bass, 1975.

Hobson vs. Hansen, 269 F. Supp. 401 (1967).

Humphreys, L., & Stubbs, J. A longitudinal analysis of teacher expectation, student expectation, and student achievement. Journal of Educational Measurement, 1977, 14, 261-270.

Hunter, J. & Schmidt, F. Critical analysis of the statistical and ethical implications of various definitions of test bias. Psychological Bulletin, 1976, 83, 1053-1071.

Jackson, G. On the report of the ad hoc committee on educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 88-92.

Jensen, A. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Kamin, L. The science and politics of IQ. Potomic, MD: Erlbaum Associates, 1974.

Kaufman, A. & Doppelt, J. Analysis of WISC-R standarization data in terms of stratification variables. Child Development, 1976, 47, 165-171.

Kolstoe, O. Teaching educable mentally retarded children (2nd ed.). New York: Holt, 1976.

Lambert, N., Windmiller, M., Cole, L. & Figueroa, R. Manual: AAMD adaptive behavior scale, public school version (1974 rev.). Washington, DC: American Association on Mental Deficiency.

Lamprecht, M. & Reschly, D. Expectancy effects of labels: Fact or artifact? Unpublished paper, 1978.

Larry P. et. al. vs. Wilson Riles et. al. United States District Court, Northern District of California, Case No. C-71-2270 RFP.

Lau vs. Nichols. 414 U. S. Pp. 563-572, 1974 (Also see Oakland, 1977).

Lesser, G., Fifer, G. & Clark, D. Mental abilities of children from different social-class and cultural group. Monographs of the Society for Research in Child Development, 1965, 30(4), 1-115.

Loehlin, J., Lindzey, G., & Spuhler, J. Race differences in intelligence. San Francisco: Freeman, 1975.

MacMillan, D. Mental retardation in school and society. Boston: Little, Brown, 1977.

MacMillan, D., Jones, R., & Aloia, G. The mentally retarded label: A theoretical analysis and review of research. American Journal of Mental Deficiency, 1974, 79, 241-261.

Mercer, J. Labeling the mentally retarded. Berkeley, CA: University of California Press, 1973.

Mercer, J. Theoretical constructs of adaptive behavior: Movement from a medical to a social-ecological perspective. In Coulter, A. & Morrow, H. (Eds.) The concept and measurement of adaptive behavior. New York: Grune and Stratton, 1978.

Mercer, J. & Lewis, J. Technical Manual: SOMPA. New York: Psychological Corporation, 1978.

Meyers, C., Sundstrom, P., & Yoshida, R. The school psychologist and assessment in special education: A report of the Ad Hoc Committee of APA Division 16. Monographs of Division 16 of the American Psychological Association, 1974, 2(1), 3-57.

Mills vs. Board of Education of the District of Columbia, No. 1939-71 (D.D.C. 1971).

Novick, M. & Peterson, N. Towards equalizing educational and employment opportunity. Journal of Educational Measurement, 1976, 13, 77-88.

Oakland, T. (Ed.) Psychological and educational assessment of minority children. New York: Brunner-Mazel, 1977.

Oakland, T. Problems in the assessment of the mentally retarded and learning disabled. In Phye, G. & Reschly, D. (Eds.) School psychology: Perspectives and issues. New York: Academic Press, in press.

Oakland, T. & Matuszek, P. Using tests in nondiscriminatory assessment. In With Bias Toward None, University of Kentucky, CORRC, Lexington, KY., 1976 - Also in Oakland, T. see above.

Oakland, T. Pluralistic norms and estimated learning potential. Annual Convention of the American Psychological Association, August, 1977.

Office of Civil Rights Memorandum to State and Local Education Agencies. Dated November 28, 1972. (In Oakland, see above).

PARC - Pennsylvania Association for Retarded Children vs. Commonwealth of Pennsylvania, 334 F. Supp. 1257 (E.D. Pa. 1971)

Peterson, N. & Novick, M. An evaluation of some models for culture fair selection. Journal of Educational Measurement, 1976, 13, 3-29.

Rapaport, D., Gill, M. & Schafer, R. Diagnostic psychological testing (Revised edition by R. Holt). New York: International Universities Press, 1968.

Rehabilitation Act of 1973, Section 504, Federal Law 93-380.

Reschly, D. Sociocultural influences on level and pattern of measured intelligence for four groups. Unpublished manuscript, 1978.

Reschly, D. WISC-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. Journal of Consulting and Clinical Psychology, 1978, in press.

Reschly, D. The school psychologist and culturally different children. In G. Phye & D. Reschly (eds.) School Psychology: Perspectives and Issues. New York: Academic Press, 1978 (in press).

Reschly, D. & Reschly, J. Predictive utility of WISC-R factor scores for four groups. Journal of School Psychology, in press.

Reschly, D. & Sabers, D. An examination of test bias for Black, Chicano, and Native American Papago Children. Unpublished manuscript, 1978.

Reschly, D. & Jipson, F. Ethnicity, geographic locale, age, sex, and urban-rural residence as variables in the prevalence of mild retardation. American Journal of Mental Deficiency, 1976, 81, 154-161.

Rosenthal, R. & Jacobsen, L. Pygmalion in the classroom: Teacher expectations and pupils' intellectual development. New York: Holt, 1968.

Rudman, H. The standardized test flap. Kappan, 1977, 59, 179-188.

Samuda, R. Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead and Co., 1975.

Sattler, J. Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73, 137-160.

Sattler, J. Intelligence testing of ethnic minority-group and culturally disadvantaged children. In Mann, L. and Sabatino, D. (Eds.) The first review of special education. Vol. 2. Philadelphia: The JSE Press, 1973.

Sattler, J. Assessment of children's intelligence. Philadelphia: W. B. Saunders, 1974.

Shockley, W. Models, mathematics, and the moral obligation to diagnose the origin of Negroe IQ deficits. Review of Educational Research, 1971, 41, 369-377.

Trotman, F. Race, IQ, and the middle class. Journal of Educational Psychology, 1977, 69, 266-273.

Tucker, J. Operationalizing the diagnostic-intervention process. In T. Oakland (Ed.) Psychological and Educational Assessment of Minority Children. New York: Brunner/Mazel, 1977.

Wade, T. & Baker, T. Opinions and use of psychological tests: A survey of clinical psychologists. American Psychologist, 1977, 32, 874-882.

Williams, R. Abuses and misuses in testing black children. The Counseling Psychologist, 1971, 2, 62-73.

Williams, R. The problem of the match and mismatch in testing black children. In L. Miller (Ed.), The Testing of Black Students: A Symposium. Englewood Cliffs, NJ: Prentice-Hall, 1974.

Williams, R. The BITCH-100: A culture-specific test. Journal of Afro-American issues, 1975, 3, 103-116.

Yoshida, R., MacMillan, D. & Meyers, E. The decertification of minority group EMR students in California: Student achievement and adjustment. In Jones, R. (Ed.) Mainstreaming and the minority child. Reston, VA: Council for Exceptional Children, 1976.

Yoshida, R. & Meyers, E. Effects of labeling as Educable Mentally Retarded on teachers' expectancies for change in a students' performance. Journal of Educational Psychology, 1975, 67, 521-527.

Ziskin, J. Coping with psychiatric and psychological testimony. Beverly Hills: Law and Psychology Press, 1975.

APPENDICES

APPENDIX A. GUADALUPE COURT DECISION

JERRY LEVINE
DIRECTOR OF LITIGATION
MARICOPA COUNTY LEGAL AID SOCIETY
P. O. Box 3076
Tempe, Arizona 85281
(602) 966-1138
Attorney for Plaintiffs

STUART R. ABELSON
STAFF ATTORNEY
HARVARD UNIVERSITY CENTER FOR
LAW AND EDUCATION
38 Kirkland Street
Cambridge, Massachusetts 02138
(617) 495-4666
Co-counsel

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF ARIZONA

GUADALUPE ORGANIZATION, et. al., Plaintiffs, - vs - TEMPE ELEMENTARY SCHOOL DISTRICT NO. 3, et. al., Defendants.	NO. CIV 71-435 PHX. STIPULATION AND ORDER

IT IS HEREBY STIPULATED by the parties, through their attorneys undersigned,

1. That acting pursuant to the Stipulation of September 9, 1971, certain of the parties to the above-captioned matter,

to wit: Plaintiffs' attorney and State Defendants, namely, W. P. SHOFSTALL, Superintendent, Arizona Department of Education and Member, State Board of Education; DAVID WEISENBORN, President State Board of Education; JOSEPH P. RALSTON, MAURICE A. MARKS, RICHARD L. HARRIS, Members, State Board of Education; DONALD M. JOHNSON, State Director of Division of Special Education, through their attorneys, have met in good faith negotiations to resolve the problems alleged in the Complaint and have developed additional and clarifying regulations as are set forth below.

2. That the parties involved with the assistance of the "Special Committee" established in accordance with the Stipulation of September 9, 1971, have arrived at interim regulations and directives which became effective December 20, 1971. (Said interim regulations are stated in #4 below.);

3. That State Defendants agree that permanent regulations which also have been developed by the method stated in #2 above and which are hereinafter set forth in #5 below shall be implemented along with all interim regulations and all other regulations, addendums, and directives currently in effect upon the next publication of the Administrator's Guide which should be on or about July 15, 1972;

4. That the interim regulations and directives mentioned above are as follows:

A. Special education directive #1, December 20, 1971.

"At the earliest practical moment, a reevaluation will be completed, in accordance with the regulations of the TENTATIVE ADMINISTRATOR'S GUIDE, Programs for Exceptional Children, 1971-72, (and addendums to the Guide), on all children whose primary language is determined to be other than English and who are presently enrolled in programs for the Educable Mentally Handicapped or Trainable Mentally Handicapped in order to assure appropriate placement."

B. Addendum to TENTATIVE ADMINISTRATOR'S GUIDE, Programs for Exceptional Children, December 20, 1971.

". . . before a child is evaluated for placement in a special education program, a determination will be made as to whether his primary language is other than English. Each school district shall follow the procedure developed by the Division of Special Education to determine a child's primary language." Said guidelines for determining primary language are on Page 1 of said addendum. "The chief school administrator will attest in writing to the investigation, and determination of the individual's primary language prior to evaluation of the student for special education. If a child's primary language is determined to be other than English, a school district shall follow one or more of the listed objectives for evaluating a child for possible placement in a special education program:

(a) Use a psychologist fluent in both the child's primary language and English;

(b) Use an interpreter to assist the psychologist both with language and testing;

(c) Use test instruments which do not stress spoken language and which are considered valid and reliable performance measures of intellectual functioning such as the Wechsler Performance Scales.

> . . . Each school district shall have results and placement of child in special education explained to parents in primary language of parents prior to placement. All information regarding any mentally handicapped status shall be privileged and confidential;

5. That permanent regulations mentioned in #3 above are as follows:

A. It is recommended that no child be placed in a special education class for the educable mentally handicapped if (a) he/she scores higher than two standard deviations below the norm on an approved verbal intelligence test in the primary language of the home; or (b) he/she scores higher than two standard deviations below the norm on an approved nonverbal intelligence test or on the nonverbal portion of an approved intelligence test which includes both verbal and nonverbal portions given in the primary language of the home. Intelligence tests shall not be either the exclusive or the primary screening device in considering a child for placement in classes for the handicapped.

B. No children shall be considered for placement in classes for handicapped children unless an examination of developmental history, cultural background, and school achievement substantiates other findings of educational handicap. This examination shall include estimates of adaptive behavior. Such examination of adaptive behavior shall include, but not be limited to, a visit, with the consent of the parent or guardian, to the child's home

by an appropriate professional adviser who may be a physician, psychologist, professional social worker or school nurse, and interviews of members of the child's family at their home. If the language spoken in the home is other than English, such interviews shall be conducted in the language of the home.

C. Where a school district enrolls any children of any racial linguistic or ethnic group in any class for exceptional children in substantially greater or lesser percentages than the percentages of such racial or linguistic or ethnic group in the school population of the district as a whole, such school district should be prepared to offer a compelling educational justification for such disproportionate enrollment.

D. No child who has not been evaluated for placement in accordance with A.R.S. § 15-1013, and the foregoing regulations shall be considered for placement in classes for exceptional children. On the basis of the evaluation carried out as prescribed above, provision for the child's educational needs shall be made by the school or schools concerned. To the degree possible the child shall be accomodated within the regular class system, with additional provision of such special and/or supportive services as may be required.

E. No child shall be considered for placement in classes for handicapped children unless the chief administrative official of the school district, or his designee have consulted, prior to

placement, the following persons pursuant to 15-1015(B) and (D) (1-5):

(a) A parent or guardian of the child.

(b) The school principal.

(c) A person responsible for administering or conducting special education courses in the school or school district.

(d) A teacher who currently has been instructing the child.

(e) An appropriate professional adviser.

One representative from each of the above categories shall meet together as an evaluation team to review the evaluation and placement of any child considered for placement in classes for handicapped children. The evaluation team shall recommend an appropriate educational program for all children considered for placement or placed in classes for exceptional children.

F. Parental approval, pursuant to § 15-1013(E), must be obtained in writing prior to placement of any child in classes for handicapped children. Such written permission shall be obtained on a form written in English and the primary language of the home, if other than English, and describing the nature and content of special programs offered, their prior effectiveness in benefitting children there assigned, and the rate and time of return of children to regular classes. Such form shall also contain notice in both English and the primary language of the home of the right of the parent or guardian to request a

review of the placement once each semester pursuant to § 15-1014, and thereafter, to withdraw consent for placement in classes for exceptional children.

G. All communication with parents of children considered for placement in classes for handicapped children, whether written or oral, shall be in the primary language of the home.

H. The parent or guardian (or his representative) of a child considered for placement in classes for exceptional children shall have complete access to all school records concerning his/her child upon request and shall receive the full cooperation of school officials in examining these records during regular school office hours.

I. All children whose primary language is determined to be other than English, attending or assigned to classes for the educable mentally handicapped or the trainable mentally handicapped at the time of the adoption of these regulations shall be reassigned to regular classes on or before October 1, 1973, unless assignment under these regulations has resulted in their being assigned to an alternative educational program.

J. Where an evaluation or reevaluation under these regulations reveals that a child has been misclassified or misplaced as a handicapped child due to difficulty in writing, speaking or understanding the English language, the school district shall remove all indications of such misclassification

or misplacement from the child's school records and shall also provide all appropriate remedial instruction.

6. That the regulations, directives, and addendums mentioned in #3 above are those in the 1971-72 edition of the Tentative Administrator's Guide which this Stipulation hereby incorporates by reference.

7. It is understood that the final wording of the above stated regulations shall be left to the discretion of the State Board of Education and its representatives. However, it is further understood that the full intent of all the aforementioned regulations must not be changed.

8. This agreement is and should be construed as full settlement of any and all claims against the State Defendants heretofore named.

9. The parties to this Stipulation agree that this will in no way affect any or all claims which the Plaintiffs have and which said Plaintiffs intend to prosecute against the County Defendants.

DATED this 24TH day of January, 1972.

Jerry Levine

JERRY LEVINE
MARICOPA COUNTY LEGAL AID SOCIETY
P. O. Box 3076
Tempe, Arizona 85281
Attorney for Plaintiffs

APPENDIX B. EXCERPTS FROM 94-142 RULES AND REGULATIONS

PROTECTION IN EVALUATION PROCEDURES

§ 121a.530 General.

(a) Each State educational agency shall insure that each public agency establishes and implements procedures which meet the requirements of §§ 121a.530–121a.534.

(b) Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory.

(20 U.S.C. 1412(5)(C).)

§ 121a.531 Preplacement evaluation.

Before any action is taken with respect to the initial placement of a handicapped child in a special education program, a full and individual evaluation of the child's educational needs must be conducted in accordance with the requirements of § 121a.532.

(20 U.S.C. 1412(5)(C).)

§ 121a.532 Evaluation procedures.

State and local educational agencies shall insure, at a minimum, that:

(a) Tests and other evaluation materials:

(1) Are provided and administered in the child's native language or other mode of communication, unless it is clearly not feasible to do so;

(2) Have been validated for the specific purpose for which they are used; and

(3) Are administered by trained personnel in conformance with the instructions provided by their producer;

(b) Tests and other evaluation materials include those tailored to assess specific areas of educational need and not merely those which are designed to provide a single general intelligence quotient;

(c) Tests are selected and administered so as best to ensure that when a test is administered to a child with impaired sensory, manual, or speaking skills, the test results accurately reflect the child's aptitude or achievement level or whatever other factors the test purports to measure, rather than reflecting the child's impaired sensory, manual, or speaking skills (except where those skills are the factors which the test purports to measure);

(d) No single procedure is used as the sole criterion for determining an appropriate educational program for a child; and

(e) The evaluation is made by a multidisciplinary team or group of persons, including at least one teacher or other specialist with knowledge in the area of suspected disability.

(f) The child is assessed in all areas related to the suspected disability, including, where appropriate, health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities.

(20 U.S.C. 1412(5)(C).)

Comment. Children who have a speech impairment as their primary handicap may not need a complete battery of assessments (e.g., psychological, physical, or adaptive behavior). However, a qualified speech-language pathologist would (1) evaluate each speech impaired child using procedures that are appropriate for the diagnosis and appraisal of speech and language disorders, and (2) where necessary, make referrals for additional assessments needed to make an appropriate placement decision.

§ 121a.533 Placement procedures.

(a) In interpreting evaluation data and in making placement decisions, each public agency shall:

(1) Draw upon information from a variety of sources, including aptitude and achievement tests, teacher recommendations, physical condition, social or cultural background, and adaptive behavior;

(2) Insure that information obtained from all of these sources is documented and carefully considered;

(3) Insure that the placement decision is made by a group of persons, including persons knowledgeable about the child, the meaning of the evaluation data, and the placement options; and

(4) Insure that the placement decision is made in conformity with the least restrictive environment rules in §§ 121a.550–121a.554.

(b) If a determination is made that a child is handicapped and needs special education and related services, an individualized education program must be developed for the child in accordance with §§ 121a.340–121a.349 of Subpart C.

(20 U.S.C. 1412(5)(C); 1414(a)(5).)

Comment. Paragraph (a)(1) includes a list of examples of sources that may be used by a public agency in making placement decisions. The agency would not have to use all the sources in every instance. The point of the requirement is to insure that more than one source is used in interpreting evaluation data and in making placement decisions. For example, while all of the named sources would have to be used for a child whose suspected disability is mental retardation, they would not be necessary for certain other handicapped children, such as a child who has a severe articulation disorder as his primary handicap. For such a child, the speech-language pathologist, in complying with the multisource requirement, might use (1) a standardized test of articulation, and (2) observation of the child's articulation behavior in conversational speech.

§ 121a.534 Reevaluation.

Each State and local educational agency shall insure:

(a) That each handicapped child's individualized education program is reviewed in accordance with §§ 121a.340–121a.349 of Subpart C, and

(b) That an evaluation of the child, based on procedures which meet the requirements under § 121a.532, is conducted every three years or more frequently if conditions warrant or if the child's parent or teacher requests an evaluation.

(20 U.S.C. 1412(5)(c).)

LEAST RESTRICTIVE ENVIRONMENT

§ 121a.550 General.

(a) Each State educational agency shall insure that each public agency establishes and implements procedures which meet the requirements of §§ 121a.550–121a.556.

(b) Each public agency shall insure:

(1) That to the maximum extent appropriate, handicapped children, including children in public or private institutions or other care facilities, are educated with children who are not handicapped, and

(2) That special classes, separate schooling or other removal of handicapped children from the regular educational environment occurs only when the nature or severity of the handicap is such that education in regular classes with the use of supplementary aids and services cannot be achieved satisfactorily.

(20 U.S.C. 1412(5)(B); 1414(a)(1)(C)(iv).)

§ 121a.551 Continuum of alternative placements.

(a) Each public agency shall insure that a continuum of alternative placements is available to meet the needs of handicapped children for special education and related services.

(b) The continuum required under paragraph (a) of this section must:

(1) Include the alternative placements listed in the definition of special education under § 121a.13 of Subpart A (instruction in regular classes, special classes, special schools, home instruction, and instruction in hospitals and institutions), and

(2) Make provision for supplementary services (such as resource room or itinerant instruction) to be provided in conjunction with regular class placement.

(20 U.S.C. 1412(5)(B).)

§ 121a.552 Placements.

Each public agency shall insure that:

(a) Each handicapped child's educational placement:

(1) Is determined at least annually,

(2) Is based on his or her individualized education program, and

(3) Is as close as possible to the child's home;

(b) The various alternative placements included under § 121a.551 are available to the extent necessary to implement the individualized education program for each handicapped child;

Taken from Federal Register, August 23, 1977.

APPENDIX C. OUTLINE OF NONBIASED ASSESSMENT PROCEDURES DEVELOPED BY THE NORTHEAST REGIONAL RESOURCE CENTER

REFERRAL

1. Are the parents/guardians aware that a referral has been made for their child, and by whom?

2. Is this child's presenting problem clearly and precisely stated on the referral?

    a. Does the referral include descriptive samples of behavior rather than opinions of the referring agent?

    b. Is there supportive documentation of the problem?

3. Is the referral legitimate?

    a. Does the referring agent have a history of over referral of children from certain cultural groups?

    b. Could irrelevant personal characteristics (e.g. sex or attractiveness) of the child have influenced the decision to refer him?

    c. Could the referring agent have misinterpreted this child's actions or expression due to his lack of understanding of cultural differences between himself and the child?

4. Can the assessment team provide the referring agent with interim recommendations that may eliminate the need for a comprehensive evaluation?

    a. Is it possible that the curriculum being used assumes that this child has developed readiness skills at home that in reality he hasn't had the opportunity to develop? If so, can the team assist the teacher in planning a program to give this child the opportunity to develop readiness skills?

    b. Can the team provide information on the child's cultural background for the referring agent so that there are fewer misunderstandings between the referring agent and this child and perhaps other children of similar cultural background?

5. Have I informed this child's parents/guardians in their primary language of the referral?

   a. Have I explained the reason(s) for the referral?

   b. Have I discussed with the parents what next step activities may be involved?
      e.g. - professional evaluations
      - use of collected data
      - design of an individualized educational plan, if necessary

   c. Have I discussed due process procedures with the parents?

   d. Do I have documented parental permission for the evaluation?

   e. Have I asked the parents to actively participate in all phases of the assessment process?

   f. Have I informed the parents of their right to examine all relevant records in regard to the identification, evaluation and educational plan of their child?

MEETING THE CHILD

1. What special conditions about this child do I need to consider?

   a. What is the child's primary home language?

   b. Do I know about the child's home environmental factors?
      e.g. - familial relationships/placement
      - social and cultural customs

   c. Do I understand this child's culture and language so that I can evoke a level of performance which accurately indicates the child's underlying competencies?

   d. Is this child impeded by a handicap other than the referral problem that may result in his not understanding what I am talking about?

2. What special conditions about myself do I need to consider?

   a. How do I feel about this child?

   b. Are my values different from this child's?

   c. Will my attitude unfairly affect this child's performance?

   d. Can I evaluate this child fairly and without prejudice?

   e. If not, would I refer him to another assessor if one is available?

3. Have I examined closely all the available existing information and sought additional information concerning this child?

   a. Has the child's academic performance been consistent from year to year?

   b. Is there evidence in this child's record that his performance was negatively or positively affected by his classroom placement or teacher?

   c. Are his past test scores consistent with his past class performance?

   d. Am I familiar with past test instruments used to evaluate this child and how well can I rely on his prior test scores?

   e. Have I observed this child in as many environments as possible (individual, large group, small group, play, home)?

   f. Am I making illegitimate assumptions about this child? e.g. Do I assume he speaks and reads Spanish simply because he is Puerto Rican?

3. g. Have I actively sought additional information on non-school related variables that may have affected this child's school performance?

e.g. - health factors (adequate sleep, food)
- family difficulties
- peer group pressures

4. Does this child understand why he is in the assessment situation?

a. Have I tried to explain at his level of understanding what the reasons were for his referral?

b. Have I given this child the opportunity to freely express his perceptions of "the problem"?

c. Have I discussed with the child what next step activities may be involved?

SELECTION OF APPROACH FOR ASSESSMENT

1. Have I considered what the best assessment approach is for this child?

a. Considering the reasons for referral, do I need to utilize behavioral observations, interviews, informal techniques or standardized techniques or a combination of the above?

b. Have I given as much thought to assessing this child's adaptive behavior as I have to his academic school performance?

c. Are the approaches I am considering consistent with the child's receptive and expressive abilities?

d. Am I placing an overdependence on one technique and overlooking others that may be more appropriate?

e. Have I achieved a balance between formal and informal techniques in my selection?

2. If I have selected to use standardized instruments, have I considered all of the ramifications?

   a. Am I testing this child simply because I've always used tests in my assessment procedure?

   b. Am I administering a particular test simply because it is part of THE BATTERY?

   c. Am I administering a test because I have been directed to do so by the Administration?

   d. Does the instrument I've chosen include persons in the standardization sample from this child's cultural group?

   e. Are subgroup scores reported in the manual?

   f. Were there large enough numbers of this child's cultural group in the test sample for me to have any reliance on the norms?

   g. Does the instrument I have selected assume a universal set of experiences for all children?

   h. Does the instrument selected contain illustrations that are misleading and/or outdated?

   i. Does the instrument selected employ vocabulary that is colloquial, regional and/or archaic?

   j. Do I understand the theoretical basis of the instrument?

   k. Will this instrument easily assist in delineating a recommended course of action to benefit this child?

   l. Have I reviewed current literature regarding this instrument?

   m. Have I reviewed current research related to potential cultural influences on test results?

TEST ADMINISTRATION

1. Are there factors (attitude, physical conditions) which support the need to reschedule this child for evaluation at another time?

2. Could the physical environment of the test setting adversely affect this child's performance?

   - room temperature
   - noise
   - inadequate space
   - poor lighting
   - furnishings inappropriate for child's size

3. Am I familiar with the test manual and have I followed its directions?

4. Have I given this child clear directions?

   a. If his native language is not English, have I instructed him in his language?

   b. Am I sure that this child understands my directions?

5. Have I accurately recorded entire responses to test items, even though the child's answers may be incorrect, so that I might later consider them when interpreting his test scores?

6. Did I establish and maintain rapport with this child throughout the evaluation session?

SCORING AND INTERPRETATION

1. Have I examined each item missed by this child rather than merely looking at his total score?

   a. Is there a pattern to the types of items this child missed?

   b. Are the items missed free of cultural bias?

   c. If I omitted all items missed that are culturally biased, would this child have performed significantly better?

2. Am I aware that I must consider other factors in the interpretation of this child's scores?

   a. Have I considered the effect the child's attitude and/or physical condition may have had on his performance?

   b. Have I considered the effect that the child's lack of rapport with me may have had on his performance?

   c. Does my interpretation of this child's performance include observations?

   d. Do I realize that I should report and interpret scores within a range rather than as a number?

3. What confidence do I have in this child's test scores?

   a. Are test scores the most important aspect of this child's evaluation?

   b. Will I allow test scores to outweigh my professional judgement about this child?

CONSULTATION WITH TEAM MEMBERS AND OTHERS

1. Am I working as an integral member of a multidisciplinary team on behalf of this child?

   a. Have I met with the team to share my findings regarding this child?

   b. Are other team member's evaluation results in conflict with mine?

   c. Can I admit my discipline's limitations and seek assistance from other team members?

   d. Do I willingly share my competencies and knowledge with other team members for the benefit of this child?

e. Has the team arrived at its conclusions as a result of team consensus or was our decision influenced by the personality and/or power of an individual team member?

2. Is the multidisciplinary team aware of its limitations?

a. Are we aware of community resource personnel and agencies that might assist us in developing an educational plan for this child? Do we utilize such resources before, during, and after the evaluation?

b. Do we on the team feel comfortable in including this child's parents in our discussions?

ASSESSMENT REPORT

1. Is my report clearly written and free of jargon so that it can be easily understood by this child, his parents, and teachers?

2. Does my report answer the questions asked in the referral?

3. Are the recommendations I have made realistic and practical for the child, school, teacher and parents?

4. Have I provided alternative recommendations?

5. Have I included in my report a description of any problems that I encountered and the effects of such during the assessment process?

INDIVIDUAL EDUCATIONAL PLAN

1. Are we making this child fit into an established program or are we developing an individualized educational plan appropriate for this child?

a. Have we identified this child's strengths and weaknesses?

b. Have we specified long range goals and immediate objectives for this child?

c. Are we willing to assist the teacher in implementing this child's educational plan?

d. Have we stated when and how this child's progress will be evaluated and by whom?

FOLLOW UP

1. What are my responsibilities after we have written this child's educational plan?

   a. Have I discussed my findings and recommendations with this child's parents and explained their due process rights? have I given the parents a written copy of this child's educational plan?

   b. Have I met with those working with this child to discuss the educational plan and to assist them in implementing its recommendations?

   c. Have I discussed my findings and recommendations with this child at his level of understanding?

   d. Can I help those working directly with the child to become more familiar with this child's social and cultural background?

   e. Have I sought this child's parents' permission for release of any confidential materials to other agencies and professionals?

   f. Will I periodically review this child's educational plan in regard to his actual progress so that any necessary changes can be made?

SOME FINAL THOUGHTS

1. Do I believe in the right to an appropriate education for all children?

2. Would I be comfortable if MY child had been involved in THIS assessment process?

3. Is there a willingness and desire on my part to actively participate in in-service activities that will lead to the further development of my personal and professional growth?

APPENDIX D.[1]

DRAFT NO. 3
SPECIAL EDUCATION ASSESSMENT PROCEDURES
DPI Committee on Nonbiased Assessment
(Dan Reschly, George Garcia, Jeff Grimes, Wilbur House, Merry Maitre, Pat O'Rourke, and Wayne Mooers)

I. PROGRAMMING AND INTERVENTION IN THE REGULAR CLASSROOM.

A. Basic Principle: Prior to referral to special education diagnostic services, solutions to classroom learning and adjustment problems should be attempted in the regular classroom.

B. Basic Principle: Various resource personnel, e.g., remedial reading specialists, curriculum consultants, counselors, psychologists, speech clinicians and social workers, should be available to assist teachers in developing educational procedures for meeting the child's needs in the regular classroom.

Considerations:

1. Are specially trained personnel available to assist classroom teachers and do these personnel provide assistance to teachers in developing alternative procedures in the regular classroom?

2. What changes are made in the regular classroom programs in order to serve children with diverse backgrounds and diverse characteristics?

3. What alternative materials and approaches, independent of special education, exist and have been attempted for children with learning and adjustment problems?

4. In cases, referred to special education services what evidence exists to confirm that attempts were made to solve the problem within the regular classroom? Were special personnel involved? Was an organized plan developed? Was the plan implemented? Was the plan given sufficient time to be successful?

5. Were efforts made to inform parents of the problem and attempted solutions, and were parents given an opportunity to contribute to solutions attempted in the regular classroom?

II. SCREENING AND REFEREAL PHASE.

A. Basic Principle: Prior to formal diagnostic procedures, adequate information should be obtained which establishes the nature and extent of deviation from reasonable expectations.

[1]This document reflects the current thinking of the committee. The document has not been approved officially by any division of DPI, and may be changed as a result of discussions with DPI and AEA personnel.

Considerations:

1. Is the concern related to classroom learning or adjustment stated or restated specifically in behavioral terms rather than in terms of a special education category?

2. Is the concern related to current classroom learning or adjustment supported and illustrated by descriptive samples of behaviors?

3. Is consideration given to and evidence provided concerning the child's strengths within school and in other situations?

4. Are other sources of information considered systematically? Is this information consistent or inconsistent with the referral? Other sources of information should include the educational history (evaluations by previous teachers, previous educational methods and materials used, previous grades), achievement test scores, previous evaluations by support personnel, previous and current social and emotional patterns of behavior, etc.

5. Do the above sources of information confirm the need for consideration of special education alternatives or does the information suggest that solutions should be attempted within the regular classroom?

B. Basic Principle: Parental involvement shall be obtained in all phases of referral, evaluation, and placement. Informed consent and due process procedures should be initiated early and followed throughout. (See Iowa DPI Special Education Rules and Regulations for description of procedures.)

Considerations:

1. Are parents informed of the reasons for the referral in precise, meaningful language?

2. Have all communications been in the primary language of the home?

3. Does the school use a variety of means to solicit active parental participation in all phases of evaluation and staffing? Are parents informed of their rights to examine all relevant records?

4. Are parents provided with information concerning the activities and kind of decisions anticipated in evaluation and staffing along with estimates of time required, and specification of personnel responsible?

III. EVALUATION

A. Basic Principle: The evaluation of children referred for special education services should be conducted by a multidisciplinary team.

Considerations:

1. Is someone assigned the responsibility of coordinating the work of the team members including, a) evaluating the referral, b) determining the kind of information needed, c) assigning appropriately trained personnel to collect the data, d) facilitating communication among the team members?

2. Are interim procedures established for assisting the child and classroom teacher while the evaluation and staffing are conducted?

B. Basic Principle: Multifactored Assessment. Children should be assessed in all areas related to the suspected handicap including where appropriate health, vision, hearing, adaptive behavior, sociocultural background, emotional status, academic performance, aptitude (intelligence), language, and psychomotor. No single procedure such as IQ test results is used as the primary source of information, and the assessment procedures are used to identify areas of special educational needs. "Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory." (Public Law 94-142, Section 121a 530, Part b.)

Considerations:

1. Situational Assessment. Is an assessment of the school or classroom environment conducted which includes a behavioral definition of the referral problems? Are data collected on the frequency and magnitude of the problem(s), and a study made of the anticedent, situational, and consequent conditions related to the problem?

2. Health History: Are data collected on physical/health conditions which may be related to the learning problem? This information would include factors such as developmental history, disease and injury data, sensory status, medications(s) used, and nuitrution?

3. Personal and Social Adjustment. Is personal and social adjustment (adaptive behaviors) in the home, neighborhood, and broader community evaluation using formal and informal data collection procedures?

4. Personal and Social Adjustment. Is personal and social adjustment (adaptive behaviors) in the school setting evaluated with formal and informal data collection procedures?

5. Primary Language. Is the child's primary language dominance determined, and are the assessment procedures administered and interpreted in a manner consistent with the primary language data?

6. Social and Cultural Background. Is the sociocultural background of the child assessed systematically, and are the results of other assessment procedures interpreted in light of the sociocultural data?

7. Educational Achievement Norm Referenced. Is educational achievement assessed with norm referenced instruments which yield valid information concerning the child'a current performance in relation to grade level expectancies?

8. Educational Achievement Criteria Referenced. Is education achievement assessed with criterion referenced instruments or devices which provide valid information concerning specific skills and deficit areas?

9. Aptitude. Is academic aptitude, i.e., general intelligence, assessed with appropriate instruments available, consideration given to variations in performance over different factors of academic aptitude, and results interpreted in veiw of strengths and limitations of such measures?

10. Psychoeducational Process. Are psychoeducational processes and motor skills related to learning assessed, and the influence of these factors on the learning or adjustment problem considered. (e.g., attention, eye-hand coordination, language, visual-motor, visual perception, auditory discrimination, etc.)

11. Other Information. Is information from other areas of assessment potentially important to placement and educational programming considered, e.g., career and vocational interests and aptitudes?

IV. STAFFING

A. Basic Principles: Placement decisions should be based upon information from a variety of sources (see previous section). Consideration of the information from the multifactored assessment should be documented in the staffing report. Placement decisions should be made by a group of persons including appropriate professional personnal and parents. The least restrictive alternative principle shall guide the selection of option for serving children.

Considerations:

1. What evidence exists which documents the consideration of a broad variety of information, including both strengths and deficits, in determining educational needs and selection of placement options?

2. Does the determination of education needs and selection of placement option include the contributions of relevant professional personnel and parents?

3. Are current educational status and educational needs stated precisely and supported by data?

4. Are alternative options considered for meeting these needs including regular education with or without support services?

5. Are special education eligibility recommendations made in conformance with the criteria for primary handicapping condition as defined in the Department of Public Instruction Special Education Rules and Regulations?

6. In making the special education eligibility recommendations, did the multidisciplinary team consider a broad variety of information including adaptive behavior and sociocultural background? How did this information influence the recommendations concerning goals for intervention and placement option?

7. Are a variety of program options considered in view of the information from the multifactored assessment? For example, using information on adaptive behavior outside of school to choose between special classes and resource options for mild or minimal mental disabilities?

8. What evidence supports the choice of program option as an appropriate alternative for meeting the child's needs?

9. Is an interim plan developed and implemented to assist the child in the regular classroom until the placement recommendations are carried out?

10. Do the special education personnel inform parents of the primary handicapping condition (If any) and explain the full range of available alternatives for meeting the child's needs?

11. Do parents contribute to decisions concerning the objectives of special education services and to choices concerning type of special education service selected.

12. Are there provisions for members of the multidisciplinary staffing team to express opinions which disagree with the decision of the majority? Are the dissenting opinions in written form expressing the reasons for disagreement?